National Institute on Drug Abuse
RESEARCH
MONOGRAPH SERIES

Statistical Issues
in Clinical Trials
for Treatment of
Opiate
Dependance

128

Statistical Issues in Clinical Trials for Treatment of Opiate Dependence

Editor:

Ram B. Jain, Ph.D.

NIDA Research Monograph 128
1992

ACKNOWLEDGMENT

COPYRIGHT STATUS

Contents

Introduction

Ram B. Jain

The Medications Development Division (MDD) of the National Institute on Drug Abuse (NIDA) came into existence in August 1990. Its mandate from the U.S. Congress is to develop medications for the treatment of drug dependence, primarily heroin and cocaine dependence. The organizational structure of MDD allows for five branches, one of which is the Biometrics Branch. I happened to be the first one to join the Biometrics Branch, and it was and still is a great learning opportunity for me. I found: Drug dependence is not a disease in the traditional sense that cancer or heart disease is; its treatment is not a treatment in the traditional sense-drug dependence is not treated the way a cancer or an infection is treated; and the characteristics of the data generated by clinical studies in drug abuse area are unique, not seen in other branches of medicine-a more than 50-percent dropout rate! The data generated by these studies are the product of a continuous dynamic interaction between the pharmacological effect of the therapeutic agent, the effect of nonpharmacological services provided as part of the total treatment, and most importantly, the drug-seeking behavior of the addict, which is shaped and influenced by the environmental stimuli around him or her. How does one statistically adjust for this multidimensional "noise"? What is being treated here is not quite obvious-Is it a medical condition, a mental disorder, a behavioral abnormality, or all of them at the same time?

Between September 1988 and May 1990, Drs. Rolley E. Johnson and Paul J. Fudala conducted a randomized double blind, "double dummy" clinical trial (ARC 090) to evaluate the efficacy of 8 mg sublingual doses of buprenorphine compared with 20 mg and 60 mg oral doses of methadone in 162 patients. This study was conducted at NIDA's Addiction Research Center (ARC). These data were provided to me for analysis. The primary data consisted of binary (positive vs. negative) data points obtained by assaying the urine samples for the presence of opiates. Since the urine samples were obtained three times a week from each patient in this 25-week study, each patient could provide up to 75 data points. Many endpoints could be defined and clinically defended using these data (e.g., percent-positive samples; a drug-free period of, say, 28 days or more), and several different statistical methods could be used to analyze them. After spending several months with these data, finding

myself more informed every day than the day before, I determined that more could be learned—I could use expert opinion from outside.

During the summer of 1991, I began planning for a workshop (a NIDA technical review) in design and analysis of clinical trials in the treatment of opiate dependence. Many well-known statisticians, including those who had many years of experience in managing and analyzing clinical trials, were contacted and asked if they would like to write and present research papers on the design and analysis of clinical trials in the treatment of opiate dependence and/or participate in this workshop. Commitments were obtained for five research papers. Each paper was to present the results of analyzing a part of the ARC 090 data. I also decided to present two papers-one on design, one on analysis. The statisticians who agreed to write research papers and/or participate (and finally came to the workshop) included Drs. Joseph Collins (Veterans' Administration Medical Center), Lloyd D. Fisher (University of Washington), Dean Follmann (National Heart, Lung, and Blood Institute [NHLBI]), Nancy L. Geller (NHLBI), Albert J. Getson (Merck Sharp & Dohme), Joel B. Greenhouse (Carnegie-Mellon University), Alan J. Gross (Medical University of South Carolina), Sudhir C. Gupta (Northern Illinois University), A.S. Hedayat (University of Illinois), Nicholas P. Jewell (University of California at Berkeley), Peter A. Lachenbruch (University of California, Los Angeles), Jack C. Lee (National Institute of Child Health and Human Development [NICHD]), Mei-Ling Ting Lee (Boston University), Shou-Hua Li (National Institute of Dental Research), Taesung Park (NICHD), Carol K. Redmond (University of Pittsburgh), Saul Rosenberg (NIDA), Vincent Shu (Abbott Laboratories), Richard Stein (Food and Drug Administration [FDA]), Ram C. Tiwari (University of North Carolina), L.J. Wei (Harvard School of Public Health), T.S. Weng (FDA), and Margaret Wu (NHLBI).

Without the presence, interaction, guidance, and advice of clinicians working in the drug abuse area, talking about designing and analyzing clinical trials for treatment of drug dependence would have been an exercise in futility, and therefore we requested participation from well-known clinicians in government, industry, and academia. Those who agreed to participate (and came to the workshop) included Jack D. Blaine (NIDA), Robert J. Chiarello (NIDA), Edward J. Cone (ARC), Paul J. Fudala (University of Pennsylvania), Harold Gordon (NIDA), David A. Gorelick (ARC), Charles W. Gorodetzky (CIBA-Geigy Corporation), Charles V. Grudzinskas (NIDA), John Hyde (FDA), Donald R. Jasinski (Johns Hopkins University), Rolley E. Johnson (Johns Hopkins University), Michael Murphy (Hoechst Roussel Pharmaceutical, Inc.), Frank J. Vocci (NIDA), and Curtis Wright (FDA).

The NIDA technical review on "Statistical Issues in Clinical Trials for Treatment of Opiate Dependence" took place on December 2-3, 1991, at the Bethesda Marriott, Bethesda, MD. It consisted of four sessions: a Clinical Session, a Design Session chaired by Dr. Gross, a two-part Analysis Session chaired by Drs. Wei and Fisher, respectively, and a General Issues Session cochaired by Drs. Lachenbruch and Jack C. Lee. Drs. Vocci and Johnson presented papers during the Clinical Session; Dr. Cone (with Sandra L. Dickerson) and I presented papers during the Design Session; and Drs. Follmann (with Drs. Geller and Wu), Gross, Gupta, Mei-Ling Ting Lee, Weng, and I presented papers during the Analysis Session. All papers presented during the Design and Analysis Sessions were available for precirculation and were peer reviewed prior to the meeting. Authors were also invited to write rejoinders to referees' comments. Drs. Geller, Greenhouse, Gross, Gupta, Jewell, Jack C. Lee, Redmond, and Tiwari were the reviewers. After the authors had presented their papers, reviewers also presented their comments at the workshop. Following the reviewers' comments and rejoinders, if any, there was an open brief discussion of each paper that was presented.

Individual papers during the Clinical Session were followed by a Discussion Session. The aim of this discussion session was to have the opinion of FDA about what kind of endpoints would be adequate and/or appropriate in clinical trials for treatment of drug dependence, what statistical methods should be used to analyze the data generated from these trials, and in general, what should be the strategy used to design these trials? The discussants for this session were Drs. Hyde, Gorodetzky, Stein, and Wright.

All three Statistical Sessions concluded with a combined open/panel discussion. At each of these discussion sessions, a series of questions were presented (by NIDA) to the panels for discussion. Additional questions as appropriate were allowed to be presented by any of the participants at the workshop. The members of the Design Panel were Drs. Hedayat (chair), Getson, Gross, Gupta, Jasinski, Mei-Ling Ting Lee, Redmond, and Wu. The members of the Analysis Panel were Drs. Redmond (chair), Fisher, Follmann, Greenhouse, Gross, and Hedayat. The members of the General Issues Panel were Drs. Lachenbruch (cochair), Jack C. Lee (cochair), Collins, Fisher, Gupta, Jewell, Murphy, Shu, and Tiwari.

I was honored to organize and be a participant in this NIDA technical review. The workshop was a tremendous success. There was a free exchange of opinion and information between the statisticians and clinicians. There were more agreements than disagreements. There was a unanimous agreement: These trials need a lot more work in both the design and analysis areas. However, in the unbiased opinion of a very prominent statistician, not

connected with NIDA in any way to the best of my knowledge, one of the papers presented at this workshop was what might be called a breakthrough.

This monograph presents the revised manuscripts as provided by the authors. Some of the revisions in these manuscripts may be a direct result of referees' comments and authors' rejoinders. Consequently, except for two papers, referees' comments and/or authors' rejoinders are not being reproduced, but all the referees have been given credit for their comments. Dr. Tiwari, who reviewed Dr. Gupta's paper, showed interest (after the workshop) in writing a paper. His paper is also included in this monograph. However, Dr. Gupta could not submit an acceptable revised manuscript in time for publication of this monograph. Consequently, his manuscript could not be included in the monograph.

Summaries of discussions on individual papers presented in the statistical sessions are also presented. Dr. Gross prepared the summary of discussions that followed the papers by Drs. Mei-Ling Ting Lee and Weng. I prepared all other summaries. I also prepared the summaries for the discussion session that took place during the Clinical Session and for the open/panel discussions during the Statistical Sessions. I have tried to give credit to individual speakers/ participants to the best of my ability. I have tried to reproduce opinions as close to the those of individual speakers as possible. I have tried not to inject my own biases to the degree I could. However, I take responsibility for all errors and omissions and tender my apologies to those whom I may have misrepresented and/or offended.

This is just a beginning. NIDA's MDD is busy planning the development of or is in the process of developing a variety of medications for the treatment of cocaine, heroin, and other substances that have the potential for abuse. In addition to buprenorphine (to treat heroin abuse), for which a multicentered pivotal trial is ongoing, a trial for I-alpha-acetylmethadol (LAAM) (to treat heroin abuse) will soon be initiated. This LAAM trial should lead to approval for its marketing by FDA sometime in late 1992 or early 1993. A pivotal trial for a sustained release formulation of naltrexone should be under way sometime in 1993. There are definite plans for developing a combination formulation of buprenorphine and naltrexone. New compounds are being acquisitioned from industry and elsewhere and are being tested for their potential for treatment of drug abuse.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician

Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Drug Dependence (Addiction) and Its Treatment

Frank J. Vocci, Jerome H. Jaffe, and Ram B. Jain

INTRODUCTION AND SOME DEFINITIONS

What is drug dependence or drug addiction? How does one become an addict or dependent on a drug? There is no simple or single answer to these questions. Dependence and addiction are the terms often used synonymously (as they are in this chapter). Unfortunately, these terms are often used in different ways in different contexts. Furthermore, according to Jaffe (1992):

> . . . science has been given no exclusive right to the use
> of [these] terms . . . . Among the many behaviors that have
> been labeled "addictions" in the mass media are: eating salt;
> buying lottery tickets; using gasoline, computers, or foreign
> capital; taking educational courses; watching television;
> running; and engaging in sex. Some of the uses of the term
> are deliberately metaphorical.

This chapter, however, attempts to summarize how dependence or addiction is currently viewed by most psychiatrists, physicians, and many behavioral psychologists.

Although the concept of dependence has historically been divided into *psychological dependence* and *physical (physiological) dependence,* the current approach recognizes that such terms tend to contribute to an unscientific dualism. Today, most researchers believe that the mind does not exist independently of the brain. Drug dependence involves body, brain, and behavior as influenced by the environment. Abuse of drugs does not necessarily constitute *drug dependence.* One may keep abusing a drug but may never be dependent on it and may never *need* to take it to feel normal. For someone to change from being a drug abuser who is nondependent to someone who is drug dependent, the sense of control must change so that the individual begins to feel a need to take the drug to feel normal, and therefore the flexibility to use or not to use the drug is diminished. During this transition

from nondependeny to dependency, the pattern of drug abuse does not have to change, although quite often there is an escalation in terms of the number of times the drug is used or the amount that is used.

"Drug tolerance is a state of decreased responsiveness to the pharmacological effect of a drug resulting from a prior exposure to that drug or a related drug. When exposure to drug A produces tolerance to it and also to drug B, the organism is said to be cross-tolerant to drug B" (Goldstein et al. 1974). Drug tolerance can occur because of alterations in the central nervous system or because of more rapid metabolism (usually by hepatic induction). Although still used, "physical dependence" is another term that conveys a sense of sharp distinction between the brain and the "mind." Physical dependence is used to mean that the use of a given drug has produced an altered body physiology so that, when the drug is stopped, there are physiological abnormalities (which eventually pass) that can be prevented by continued use of the drug. Physical dependence can be revealed by stopping the drug or by giving an antagonist that displaces the drug from its site of action in the body. Physical dependence can result from the therapeutic uses of a drug, for example, by using opioids to relieve pain in cancer therapy or benzodiazepines to treat anxiety. The discontinuation of a drug that one is physically dependent on can result in various pathophysiologic disturbances collectively known as a withdrawal or abstinence syndrome. It is entirely possible that an individual could be physically *dependent* on a drug but still not be "addicted" to a drug; that is, the appearance of withdrawal symptoms does not necessarily cause the individual to continue using the drug. Then, what is *drug dependence?* According to Goldstein and colleagues (1974), drug *dependence* consists of three distinct and independent components: tolerance, physical dependence, and drug-seeking behavior resulting in compulsive abuse (psychic craving). Of course, these features *are* noticed in different degrees in drug dependence on different drugs. In the case of some drugs, only one or two of these components are noticed. "An example of tolerance and physical dependence without compulsive abuse is provided by the morphine congener and antagonist nalorphine" (Goldstein et al. 1974).

According to earlier concepts formulated in the 1930s, 1940s, and 1950s, a drug was not considered to be addictive unless it produced physical dependence characterized by an easily observable withdrawal syndrome. This view led to popular misconceptions about the dependence potential of both nicotine and cocaine. However, addiction is still an evolving concept. Currently, many researchers and clinicians believe that life-threatening intensity or easy observability of a withdrawal syndrome is not a necessary element in addiction. For example, nicotine is believed to be addicting even though its withdrawal syndrome is not dramatic and no one has ever died from its

withdrawal. An increasing trend in the diagnosis of dependence is to characterize the addictive disorders in terms of the pattern of use, loss of control over amounts ingested, and continued use despite medical, legal, occupational, or interpersonal problems.

There are now two widely recognized sets of standard criteria that are used to determine whether a given individual should be considered to be dependent on a drug: the DSM-III-R criteria developed by the American Psychiatric Association (1987) and the ICD-10 criteria developed by the World Health Organization (1990). The DSM-III-R criteria for drug dependence include behaviors that allow an observer to infer that the individual has a decreased freedom to choose whether or not to use the drug. To be diagnosed as drug dependent, a person must meet three of the following criteria (American Psychiatric Association 1987):

- Ingestion of larger amounts (of drug) or over a longer period of time than intended, signifying loss of control over behavior

- Desire to or unsuccessful attempt to cut down drug use, once again representing loss of control over behavior

- Great deal of time spent in procuring drug and recovering from its effects

- Frequent intoxication or withdrawal when expected to fulfill major role obligations at work, school, or home; i.e., interference with obligations of life; e.g., reinforcing things in life like watching TV, reading books, interactions with people etc.

- Other activities given up or reduced due to substance use

- Continued use despite problems at work, in life (e.g., marital problems) or legal problems

- Marked tolerance

- Characteristic withdrawal symptoms

- Substance use to relieve withdrawal

In addition, these symptoms or behaviors must persist for more than 1 month. Furthermore, drug dependence can be graded as mild, moderate, or severe depending on the number of criteria met. A full remission means no use or use with no dependence in the past 6 months.

The criteria used in ICD-10 are somewhat different. According to ICD-10, for someone to be diagnosed as (drug) dependent, at least three of the following should have been experienced or exhibited at some time during the previous year (World Health Organization 1990):

- A strong desire or sense of compulsion to take the substance

- An impaired capacity to control substance taking behavior in terms of onset, termination or levels of use

- Substance use with intention of relieving withdrawal symptoms and with awareness that this strategy is effective

- Physiological withdrawal state

- Evidence of tolerance such that increased doses of the substance are required in order to achieve effects originally produced by lower doses

- Narrowing of the personal repertoire of patterns of substance use

- Progressive neglect of alternative pleasures or interests in favor of substance use

- Persisting with substance use despite clear evidence of overly harmful consequences

However, neither of these sets of criteria is used by the Federal Government for admission to a methadone maintenance program. According to Federal regulations, dependence criteria for admission to a methadone maintenance program are at least 1 year of addiction history, physiological addiction for at least 1 year, andcontinuous or episodic addiction for most of the preceding year (Methadone maintenance criteria 1989). It would be inappropriate to view this as a formal definition of addiction; rather, it should be seen as specifying a degree of addiction or opioid dependence that justifies admission to a

specialized program. In one sense, however, one could say that there is no standard definition of drug dependence or any standard diagnostic test that can be administered to classify a drug-dependent individual in need of treatment. However, in the case of opioid dependence, there is a naloxone challenge test that, by displacing opioids from the receptors in the brain, will produce signs of physical dependence, that is, withdrawal symptoms, in anyone who has been using opioids for a few days or longer. This test can also be given to an individual who might be taking opioids for therapeutic purposes (and will produce the same withdrawal symptoms after even a few doses of opioids). Hence, the presence of a withdrawal syndrome (even a severe one) does not necessarily mean the individual is addicted. The presence of a withdrawal syndrome is neither necessary nor a sufficient condition for the diagnosis of drug dependence. However, as noted above, in an individual with a history of abuse, the presence of a withdrawal syndrome should be documented when that person is seeking admission to a methadone maintenance program. Hence, for the purpose of a clinical trial, the definition (DSM-III-R or ICD-10) of dependence with or without additional criteria (e.g., naloxone challenge scores) can be used. Using DSM-III-R criteria allows entrance into clinical trials of patients who would not necessarily meet criteria for admission to a methadone maintenance program.

TREATMENT OF OPIOID ADDICTION

There are more than 1 million opioid abusers in the United States who can possibly benefit from a treatment program. Of these, about 110,000 are in methadone maintenance programs, and about 3,000 are in naltrexone treatment. Many others are treated in detoxification programs, therapeutic communities, and 12-step, drug-free programs; it is likely that the overwhelming majority of this population are not participating in any kind of treatment.

Although pharmacologically based treatments are only one approach to treatment, this approach plays an important role in the American system. There are primarily two pharmacological approaches to treatment of opioid dependence: agonist therapy and antagonist therapy.

Agonist therapy for opioid dependence constitutes *replacing* the abused opioid with another, most likely a synthetic, opioid (called an opioid agonist or partial agonist) with relatively less potential for abuse. The ideal *replacement* opioid should have a less intense or no euphoric effect, should have a longer *pharmacological* effect, and should have a withdrawal effect less severe than that of the abused opioid. Replacement (maintenance) therapy may last indefinitely, although in many treatment programs the ultimate goal is to remove the addicts from all drugs and opioids.

Antagonist therapy for opioid addiction treats addicts with an opioid antagonist that blocks binding of opioids to its receptors and thus blocks *all* effects of external opioids and, perhaps in some cases, the action of endogenous opioid peptides. However, this therapy is likely to be successful only for those who are extremely *motivated* to stop using opioids or to comply with taking an antagonist (e.g., physicians who may risk losing their license to practice if they are not off the drug), In addition, the currently available antagonist agent naltrexone is not well liked by addicts for several reasons. In some individuals, it may produce negative mood states. However, these adverse effects are not usually seen in individuals who have not been dependent on opioids. However, in many cases, unwillingness to take the antagonist may stem from its therapeutic effects-it blocks the effects of opioid agonists,

As noted above, in addition to agonist and antagonist therapy, there are drug-free programs. The relapse rates for addicts who enter these programs are very high, but for small percentages who remain in TCs for 6 months or more, the outcome is generally quite positive (Vaillant 1992).

OPIOID AGONIST THERAPY

Currently, the only Food and Drug Administration (FDA)-approved pharmacotherapeutic opioid agonist for drug dependence is *methadone maintenance with counseling.* Methadone, given orally once a day to a tolerant individual, has no or little euphoric effect. Its pharmacological effect lasts for about 24 hours (thus, need for methadone arises about every 24 hours), and it has less severe though longer lasting withdrawal symptoms than heroin.

Methadone has been found to be an effective treatment in reducing the use of illicit opioids that are generally administered through an intravenous (IV) route. Since IV use and sharing of injection equipment have been associated with the spread of human immunodeficiency virus (HIV) infection, reduction in heroin IV use indirectly reduces the risk of HIV infection. Although a decrease in heroin use is seen within days after methadone is started, in opioid maintenance with methadone treatment, patients must be *stabilized* on methadone for a certain length of time before they can draw maximum benefits from the treatment. Compared with drug-free programs, considerably higher retention rates are seen in methadone treatment.

It must be mentioned here that by FDA regulation, *methadone maintenance* treatment must include other services such as counseling in addition to the administration of oral methadone. Hence, there are nonpharmacological aspects of methadone maintenance treatment. These additional services aid

addicts, for example, after they have *stabilized* to the point of ceasing to participate in crime-related activities, improving social and family relationships, and remaining in rehabilitation. The quality and quantity of these services can powerfully affect the results of treatment. Although research has shown that doses of methadone above 60 mg are more effective than lower doses in reducing heroin use, there are substantial variations in the methadone dose (10 mg per day to as much as 100 mg per day) administered in different clinics as well as in the quality and quantity of nonpharmacological services. Hence, success rates in reducing IV heroin use vary greatly from one clinic to another (Ball and Ross 1991; D'Aunno and Vaughn 1992). On the average, over a 1-month period, on a 10-mg daily dose, four of five addicts continue using heroin; on a 20 to 40 mg per day dose, about half the addicts still use heroin; on a 40 to 60 mg per day dose, only one of five addicts will use heroin; and on more than 60 mg per day doses, fewer than one in five addicts continue to use heroin, provided other services are of high quality.

However, methadone treatment is not without problems. Methadone has a protracted withdrawal, and, therefore, it is difficult to withdraw from methadone. It follows that it would be desirable to have an alternative opioid agonist that induces less severe physical dependence and from which it is easier to withdraw. Methadone is a full agonist, and fatal accidental overdoses in unintended users (e.g., nontolerant drug users, children) have been reported. A treatment agent with less toxicity would be an advantage. Methadone must be used every day, which can be costly and time-consuming and hinders rehabilitation; alternatively, addicts must be allowed take-home doses. Take-home privileges have resulted in diversion of methadone into illicit markets and, according to isolated reported cases, in the creation of methadone addicts. Hence, an agent that has longer pharmacological action (e.g., can be used twice or thrice a week rather than every day and is less susceptible to diversion) would be an advance. In addition, in certain neighborhoods and communities, methadone is not well accepted and has been perceived as a stigma. Alternative treatments that are more acceptable to such communities would be an advantage.

REFERENCES

American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders.* 3d ed., revised. Washington, DC: American Psychiatric Association, 1987.
Ball, J.C., and Ross, A. *The Effectiveness of Methadone Maintenance Treatment.* New York: Springer-Verlag, 1991.
D'Aunno, T., and Vaughn, T.E. Variations in methadone treatment practices results from a national study. *JAMA* 267:253-258, 1992.

Goldstein, A.; Aronow, L.; and Kalman, S.M. *Principles of Drug Action: The Basis of Pharmacology.* New York: Wiley, 1974. 854 pp.

Jaffe, J.H. Current concepts of addiction. In: O'Brien, *C.* P., ed. *Addictive States.* New York: Raven Press, 1992. pp. 1-21.

Methadone maintenance criteria. *Federal Register* 54(40):8954-8971, 1989.

Vaillant, G.E. Is there a natural history of addiction? In: O'Brien, C.P., ed. *Addictive States.* New York: Raven Press, 1992. pp. 1-21.

World Health Organization. 1990 draft of chapter V: Mental and behavioral disorders. Clinical descriptions and diagnostic guidelines. *International Classification of Diseases.* 10th rev. Geneva: World Health Organization, 1990.

AUTHORS

Frank J. Vocci, Ph.D.
Deputy Director
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55

Jerome H. Jaffe, M.D.
Deputy Director
Office of Treatment Improvement
Rockwall II, 10th floor

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55

5600 Fishers Lane
Rockville, MD 20857

Background and Design of a Controlled Clinical Trial (ARC 090) for the Treatment of Opioid Dependence

Rolley E. Johnson and Paul J. Fudala

INTRODUCTION

The initial clinical abuse liability study of buprenorphine was reported by Jasinski and coworkers (1978). They noted that acute single doses of buprenorphine produced morphine-like subjective, physiologic, and behavioral effects. They also found buprenorphine to be acceptable to the addict population and to block the effects of subcutaneously administered morphine. Buprenorphine appeared to have a long duration of action similar to methadone but, unlike methadone, was associated with a limited physiologic withdrawal syndrome. In the same report, the chronic subcutaneous administration of 8 mg/day of buprenorphine was equivalent to 60 mg/day of orally given methadone for subject-reported "liking." A later study (Mello and Mendelson 1980) provided additional data regarding the potential efficacy of buprenorphine by showing that it suppressed the rate of heroin self-administration by individuals participating in a clinical laboratory study.

The relative ineffectiveness by the oral route of administration (Jasinski et al. 1982) led to studies using sublingual buprenorphine. These investigations demonstrated that sublingually given buprenorphine was two-thirds as potent as when administered subcutaneously (Jasinski et al. 1989). Subsequent studies focused on various dose-induction procedures and the appropriate dose levels for the treatment of street opioid- and methadone-dependent individuals (Jasinski et al. 1983; Reisinger 1985; Seow et al. 1986; Bickel et al. 1988*a*; Kosten and Kleber 1988).

Bickel and colleagues (1988*a*) reported that sublingual buprenorphine, 2 mg/day. was significantly less effective than 30 mg of orally administered methadone in attenuating the effects of a hydromorphone challenge. The same authors later reported that the opioid-blocking activity of buprenorphine was dose related up to 8 mg/day (Bickel et al. 1988*b*), with little apparent increase in benefit when the dosage was increased to 16 mg/day.

Still to be determined were appropriate induction and dosing schedules for a clinical comparison of buprenorphine with methadone. Thus, an inpatient trial was conducted to address these therapeutic issues. Results from this study indicated that a rapid 3-day dose-induction procedure was both effective and acceptable to study participants (Johnson et al. 1989). It was also concluded that daily dosing was probably more appropriate than alternate-day dosing (Fudala et al. 1990).

The present study was designed to meet Food and Drug Administration regulatory requirements for a well-designed, well-controlled clinical trial that could be used in support of a New Drug Application for buprenorphine. To this end, the investigators attempted to control or account for those aspects of the study that could confound the data analyses or interpretation (Hargreaves 1983) including issues such as choosing appropriate design and outcome measures, subject characteristics, attrition, blinding, and others. This chapter describes the background and design of a controlled clinical trial comparing the efficacy of buprenorphine and methadone for the short-term maintenance and detoxification of opioid addicts.

DESIGN

Patients

Inclusion criteria included the following:

1. Male or female volunteers seeking treatment for opioid dependence

2. Age 21 to 50 years

3. Length of present addiction of at least 4 months

4. At least two or more episodes of heroin use per day

5. Daily value of heroin use of $50 or greater

6. A rating of 4 or greater on a self-reported level of withdrawal scale 12 hours after the last heroin dose (0=no withdrawal, 9=worst withdrawal ever experienced)

7. Three consecutively collected daily urines, at least two of which were positive for opioids but negative for methadone

Exclusion criteria included the following:

1. Any acute or chronic medical or psychiatric condition that may have compromised an individual's ability to complete the study

2. A score of 7 or higher on the interviewer severity rating of need for psychiatric/psychological treatment on the Addiction Severity Index (ASI)

3. Clinically significant abnormalities in laboratory values

4. Alanine or aspartate aminotransferase levels greater than 99 units/L on admission

Individuals were recruited through a contract service that identified potential patients from treatment, general medical, and other facilities having contact with chronic drug abusers. This service used the Shipley Institute of Living Scale and the Hopkins Symptom Checklist 90 (Revised) to ensure that prospective study participants could read and understand both the informed consent and study questionnaires and also as aids in identifying individuals who might not be qualified for the study. The study was conducted under protocol 090 at the Addiction Research Center of the Intramural Division of the National Institute on Drug Abuse (NIDA), Baltimore, MD, using its outpatient facilities. Individuals were enrolled in the trial between September 1988 and November 1989.

Each patient gave informed consent for participation in the study. The consent forms and experimental procedures were approved by the local institutional review board in accordance with the U.S. Department of Health and Human Services guidelines for the protection of human subjects.

Methods

The study was conducted using a double-blind, double-dummy (both an oral and sublingual dosage form given), parallel groups design. One dosage form contained the assigned treatment; the other was a matching placebo. The three treatment groups were:

1. Buprenorphine, 8 mg/day sublingually (n=53)
2. Methadone 20 mg/day orally (n=55)
3. Methadone 60 mg/day orally (n=54)

The 20 mg/day dosage was chosen since one-tenth of the patients in methadone clinics were treated during the initial 3 months and longer with

this or a lesser dose (U.S. Department of Health and Human Services 1984; Allison et al. 1985). Also, it has been reported that 31 percent of patients entering methadone treatment can be successfully maintained on a dose of 20 mg/day or less for 4 weeks (Peachey and Lei 1988). The 60 mg/day dosage was chosen because it was reported as the approximate median daily dosage used in maintenance therapy (U.S. Department of Health and Human Services 1984) and one that the authors hypothesized would give results significantly better than those obtained from the 20 mg/day group. The 8 mg/day dosage of buprenorphine was selected based on previous reports indicating possible efficacy (Johnson et al. 1989; Fudala et al. 1990) and effects comparable to those seen with 40 to 60 mg/day of methadone (Jasinski et al. 1978). The working hypothesis of the study was that buprenorphine 8 mg/day and methadone 60 mg/day would be more effective than methadone 20 mg/day and that buprenorphine would be at least 80 percent as effective as methadone 60 mg/day.

The dose-induction procedure is shown in table 1. Patients were subsequently continued on their maintenance dosage through study day 120.

The study consisted of 120 days of induction/maintenance followed by 49 days of gradual dosage reduction and 11 days of placebo dosing. Patients who wished to voluntarily terminate their participation in the study or who were administratively discharged were given a 21-day methadone detoxification.

For the purposes of data analysis, the study was divided into a 17-week maintenance phase (days 1 through 119) and an 8-week detoxification phase (days 120 through 175) since the detoxification phase was considered to begin with the last maintenance dose. The gradual detoxification was carried out by decreasing each treatment group's dosage by the same percentage for a given week of the study. Although the study was designed to be carried out over 175 days (25 weeks), patient participation and data collection were extended to a total of 180 days to parallel existing Federal methadone regulations for long-term detoxification.

TABLE 1.   *ARC 090 trial: dose-induction procedure*

					Study Day					
Drug/Dosage	1	2	3	4	5	6	7	8	9	10
Buprenorphine 8 mg	2	4	8	8	8	8	8	8	8	8
Methadone 60 mg	20	30	40	50	60	60	60	60	60	60
Methadone 20 mg	20	30	30	30	30	25	25	25	25	20

Stratification

Patients were stratified into treatment groups by the following criteria:

1. Age (21 to 35 and 36 to 50 years).

2. Gender

3. Clinical Institute Narcotic Assessment scores (less than 30 and greater than or equal to 30) (Peachey and Lei 1988). These scores reflect the results of a naloxone challenge test that was given to all patients immediately prior to their receiving the first dose of study medication.

Each stratification factor had two levels for a total of eight strata. Treatment assignment was performed randomly for each stratum using a permuted block design with possible block sizes of three, six, or nine.

The naloxone challenge test was used as a stratification variable to ensure approximately equivalent levels of physical dependence between groups. Age was used since various authors have shown differences in relapse and retention rates based on a patient's age (Richman 1966; Babst et al. 1971; Brown et al. 1973). Gender differences have been reported to affect retention of patients in methadone maintenance (Hser et al. 1991) and therapeutic community treatment programs (Sansone 1980). Also, since the present study incorporated fixed-dosage regimens, potential pharmacokinetic differences due to gender were controlled by stratification.

Clinic Milieu

Thirty to sixty minutes of individual counseling per week, using a relapse prevention model, was offered but not required. Medical safety was evaluated using hematology and blood chemistry panels and urinalyses collected on study days 30, 60, 90, 120, and 180. Vital signs were recorded every 2 weeks, and urine pregnancy tests were obtained every 2 months. Patient case report forms and medical records were maintained for each participant. Observed urine samples were collected three times weekly on Monday, Wednesday, and Friday. To promote patients' compliance with the urine collection process, individuals were required to submit a sample on the day(s) following a missed, scheduled collection. However, because of potential carryover and other confounds, these samples were not analyzed. Level 1 to level 2 clinical services (Childress et al. 1991) were provided to all patients.

Treatment compliance was maximized by requiring participants to come to the clinic daily to receive medication. Individuals who missed 3 consecutive days of medication were dropped from the study, with their third missed day considered to be the last day of study participation. Every effort was made to retain individuals in the study. For example, whenever possible, medications were delivered to and data collected from patients who were incarcerated in the Baltimore metropolitan area. The last day of study participation for individuals administratively discharged or those who voluntarily terminated from the study was their actual discharge or termination date.

One, zero, and three patients, randomized to the buprenorphine and methadone 20 and 60 mg/day groups, respectively, had their dosages halved due to an inability to tolerate them. Since this was a fixed-dosage protocol, these patients were considered treatment failures effective on the first day of dosage adjustment, although data collection continued. Study staff members (except pharmacy personnel) were blind to this provision of the protocol.

Primary Dependent Variables

Three primary dependent variables were identified a priori:

1. Patient retention time in the study

2. Monday, Wednesday, and Friday urine samples negative for opioids

3. Failure to maintain drug abstinence as assessed by two consecutive Monday urine samples positive for opioids following 4 weeks of treatment

The criterion for the last variable was chosen to give patients time to stabilize in treatment and to account for the probability that patients would more likely challenge the pharmacologic blockade early in treatment. Monday urine samples were selected since it was felt that patients were more likely to use (or use more) illicit opioids on weekends. A 1-week interval between samples was chosen so that a positive result would not be due to a previous sample.

Secondary Dependent Variables

Collected within the first 7 study days were results from the following:

1. Buss-Durkee Hostility Scale

2. Diagnostic Interview Schedule

3. Early Experience Questionnaire

4. Elliot Huizinga Lifetime Events Survey

5. Eysenck Impulsivity, Venturesomeness, and Empathy Questionnaire

6. Eysenck Personality Questionnaire

7. Hopkins Symptom Checklist 90 (Revised)

8. Personality Diagnostic Questionnaire

9. ASI (also obtained at study completion or termination and 3, 6, and 12 months thereafter)

The following patient-reported data were collected daily:

1. An adjective checklist (interval scale from 0 to 9) assessing opioid withdrawal symptoms, with additional items measuring urge and need for an opioid, frequent urination, and "hooked on" and "liking" for the study medication

2. A structured questionnaire (true/false) assessing opioid withdrawal symptoms

Collected three times weekly were urine samples assayed for barbiturates, benzodiazepines, cocaine metabolite, methadone, and phencyclidine.

Data collected biweekly (patient reported) included:

1. A visual analog scale assessing "want" and "need" for an opioid and cocaine

2. A 14-item medication adverse effects questionnaire

3. Beck Depression Inventory

Collected at 30, 60, 90, and 120 days and at termination were:

1. Hematology and blood chemistry panels

2. Urinalyses

3. Vital signs

Urine Toxicology

Urine samples were assayed in triplicate using appropriate positive and negative controls, once with radioimmunoassay (Abuscreen; Roche Diagnostic Systems Inc., Montclair, NJ) and twice with enzyme-multiplied immunoassay technique (EMIT; Syva Corporation, Palo Alto, CA). A sample was considered to be positive if the amount of analyte in the sample was greater than a predetermined cutoff value (e.g., 300 ng/mL for opioids). If a sample tested negative at least twice out of the three assays, it was considered negative; otherwise, it was considered positive.

Study Medications

Buprenorphine hydrochloride was obtained from Reckitt and Colman (Hull, England) through NIDA's Research Technology Branch (Rockville, MD). Drug solutions were aseptically prepared in 30 percent ethanol (vol/vol) and stored at room temperature. All solutions were administered sublingually in a volume of 1 mL using Ped-Pod oral dispensers (SoloPak Laboratories, Franklin Park, IL). Buprenorphine solutions have been shown to be stable in these dispensers for at least 3 months. To maximize the amount of buprenorphine absorbed from the sublingual mucosa, all patients were instructed to refrain from speaking and to hold the solution under the tongue for 10 minutes. Methadone HCI (methadone hydrochloride oral concentrate USP, 10 mg/mL) and cherry flavor concentrate (Mallinckrodt Inc., St. Louis, MO) were used. A methadone HCI, 2 mg/mL solution was prepared from the concentrate and distilled water. Final methadone dosages were prepared to a volume of 30 mL using this solution in a vehicle of cherry flavor concentrate:water (1:4) containing denatonium benzoate (Bitrex; J.H. Walker and Co., Inc., Mt. Vernon, NY), 0.2 ng/mL, to mask the flavor of the solutions.

SUMMARY

This study represents the largest clinical trial reported to date that demonstrated the efficacy of buprenorphine for opioid dependence treatment (Johnson et al. 1992). Although the study design was adequate to demonstrate differences between treatment groups, there has not been a consensus regarding the most appropriate method for analyzing various outcome measures of this and similar studies. To present a comprehensive review of these methods, other chapters in this monograph focus on various analytical techniques for assessing one of these measures-urine toxicology screens-for illicit opioids.

REFERENCES

Allison, M.; Hubbard, R.L.; and Rachal, J.V. *Treatment Process in Methadone, Residential, and Outpatient Drug-Free Programs.* National Institute on Drug Abuse Treatment Research Monograph Series. DHHS Pub. No. (ADM)85-1388. Rockville, MD: U.S. Department of Health and Human Services, U.S. Public Health Service, Alcohol, Drug Abuse and Mental Health Administration, 1985.

Babst, D.V.; Chambers, C.D.; and Warner, A. Patient characteristics associated with retention in a methadone maintenance program. *Br J Addict* 66:195-204, 1971.

Bickel, W.K.; Stitzer, M.L.; Bigelow, G.E.; Liebson, I.A.; Jasinski, D.R.; and Johnson, R.E. A clinical trial with buprenorphine: Comparison with methadone in the detoxification of heroin addicts. *Clin Pharmacol Ther* 43:72-78, 1988*a*.

Bickel, W.K.; Stitzer, M.L.; Bigelow, G.E.; Liebson, I.A.; Jasinski, D.R.; and Johnson, R.E. Buprenorphine: Dose-related blockade of opioid challenge effects in opioid dependent humans. *J Pharmacol Exp Ther* 247:47-53, 1988*b.*

Brown, B.S.; DuPont, R.L.; Bass, U.F. III; Brewster, G.W.; Glendinning, S.T.; Kozel, N.J.; and Meyers, M.B. Impact of a large-scale narcotics treatment program. A six month experience. *Int J Addict* 8:49-57, 1973.

Childress, A.R.; McClellan, A.T.; Woody, G.E.; and O'Brien, C.P. Are there minimum conditions necessary for methadone maintenance to reduce intravenous drug use and AIDS risk behaviors? In: Pickens, R.W.; Leukefeld. C.G.; and Schuster, C.R., eds. *Improving Drug Abuse Treatment.* National Institute on Drug Abuse Research Monograph 106. DHHS Pub. No. (ADM)91-1754. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1991. pp. 167-177.

Fudala, P.J.; Jaffe, J.H.; Dax, E.M.; and Johnson, R.E. Use of buprenorphine in the treatment of opioid addiction. II. Physiologic and behavioral effects of daily and alternate-day administration and abrupt withdrawal. *Clin Pharmacol Ther* 47:525-534, 1990.

Hargreaves, W.A. Methadone dosage and duration for maintenance treatment. In: Cooper, J.R.; Altman, F.; Brown, B.S.; and Czechowicz, D., eds. *Research on the Treatment of Narcotic Addiction. State of the Art.* National Institute on Drug Abuse Treatment Research Monograph Series. DHHS Pub. No. (ADM)83-1281. Rockville, MD: U.S. Department of Health and Human Services, U.S. Public Health Service, Alcohol. Drug Abuse, and Mental Health Administration, 1983. pp. 19-79.

Hser, Y.; Anglin, M.D.; and Liu, Y. A survival analysis of gender and ethnic differences in responsiveness to methadone maintenance treatment. *Int J Addict* 25:1295-1315. 1991.

Jasinski, D.R.; Fudala, P.J.; and Johnson, R.E. Sublingual versus subcutaneous buprenorphine in opiate abusers. *Clin Pharmacol Ther* 45: 513-519, 1989.

Jasinski, D.R.; Haertzen, C.A.; Henningfield, J.E.; Johnson, R.E.; Makhzoumi, H.M.; and Miyasato, K. Progress report of the NIDA Addiction Research Center. In: Harris, L.S., ed. *Problems of Drug Dependence, 1981: Proceedings of the 43rd Annual Scientific Meeting, The Committee on Problems of Drug Dependence, Inc.* National Institute on Drug Abuse Research Monograph 41. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1982. pp. 42-52.

Jasinski, D.R.; Henningfield, J.E.; Hickey, J.E.; and Johnson, R.E. Progress report of the NIDA Addiction Research Center, Baltimore, Maryland, 1982. In: Harris, L.S., ed. *Problems of Drug Dependence, 1982: Proceedings of the 44th Annual Scientific Meeting, The Committee on Problems of Drug Dependence, Inc.* National Institute on Drug Abuse Research Monograph 43. DHHS Pub. No. (ADM)83-1264. Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1983. pp. 92-98.

Jasinski, D.R.; Pevnick, J.S.; and Griffith, J.D. Human pharmacology and abuse potential of the analgesic buprenorphine. *Arch Gen Psychiatry* 35:501-516, 1978.

Johnson, R.E.; Cone, E.J.; Henningfield, J.E.; and Fudala, P.J. Use of buprenorphine in the treatment of opiate addiction. I: Physiologic and behavioral effects during a rapid dose induction. *Clin Pharmacol Ther* 46:335-343, 1989.

Johnson, R.E.; Jaffe, J.H.; and Fudala, P.J. A controlled trial of buprenorphine treatment for opioid dependence. *JAMA* 267:2750-2755, 1992.

Kosten, T.R., and Kleber, H.D. Buprenorphine detoxification from opioid dependence: A pilot study. Life Sci 42:635-641, 1988.

Mello, N.K., and Mendelson, J.H. Buprenorphine suppresses heroin use by heroin addicts. *Science* 207:657-659, 1980.

Peachey, J.E., and Lei, H. Assessment of opioid dependence with naloxone. *Br J Addict* 83:193-201, 1988.

Reisinger, M. Buprenorphine as new treatment for heroin dependence. *Drug Alcohol Depend* 16:257-262, 1985.

Richman, A. Follow-up of criminal narcotic addicts. *Can Psychiatric Assoc J* 11:107-115, 1966.

Sansone, J. Retention patterns in a therapeutic community for the treatment of drug abuse. *Int J Addict* 15:711-736, 1980.

Seow, S.S.W.; Quigley, A.J.; Ilett, K.F.; Dusci, L.J.; Swensen, G.; Harrison-Stewart, A.; and Rappaport, L. Buprenorphine: A new maintenance opiate? *Med J Australia* 144:407-411, 1986.

U.S. Department of Health and Human Services. *National Summary of Narcotic Treatment Programs. Annual Report for Treatment Programs Using Methadone.* Washington, DC: Supt. of Docs., U.S. Govt. Print. Off., 1984.

ACKNOWLEDGMENTS

AUTHORS

Rolley E. Johnson, Pharm.D.
Associate Professor
Department of Psychiatry and Behavioral Sciences
The Johns Hopkins School of Medicine
Building G, Room 2725
5510 Nathan Shock Drive
Baltimore, MD 21224

Paul J. Fudala, Ph.D.
Assistant Professor
Department of Psychiatry
University of Pennsylvania School of Medicine and the Department of Veterans
  Affairs Medical Center
Building 15
University and Woodland Avenues
Philadelphia, PA 19104

Clinical Endpoints: Discussion Session

Ram B. Jain

Discussants: John Hyde, Charles Gorodetzky, Richard Stein, and Curtis Wright

The aim of this discussion session was to obtain the opinion of the U.S. Food and Drug Administration (FDA) about what kind of endpoints would be adequate and/or appropriate in clinical trials for treatment of drug dependence, what statistical methods should be used to analyze the data generated from these trials, and in general, what strategy should be used to design these trials. Drs. Hyde, Stein, and Wright represented FDA, and Dr. Gorodetzky presented the pharmaceutical industry's viewpoint because FDA policy might affect its ability to conduct clinical trials.

Dr. Wright reminded that although most funded research is exploratory in nature, generating new and exciting information on the cutting edge of science, most of the drug approval work at FDA is confirmatory in nature, calling for regulatory decisions to approve or not approve drugs. As a consequence, results obtained by applying a new mathematical technique should be backed up or linked with results obtained by a mathematical technique that is known to work. Drug approval is easy when information about a new drug is coherent and robust and there is a large effect size. The results obtained in large phase III trials—generally used to support a new drug application (NDA)—should be in coherence with the results obtained from the earlier phase I and II trials in selected and general human populations and from preclinical work on animals; they should get the same answers in all those places. The conclusions obtained from analysis of data should be robust; that is, they should not be dependent on a specific experimental design, a specific method of analysis, or the specific way a trial may have been conducted. Different trials, probably using different designs, should lead to the same conclusions, This is what Dr. Stein called clinical robustness as opposed to statistical robustness. The effect size should be relatively large.

The results of the pivotal trials should not depend on a set of assumptions made at any stage of development. Outcome variables (endpoints) selected for the pivotal trials should tap several different kinds of domains. Subjective self-reports (e.g., "How are you doing today?") should be linked or obtained in parallel with observer rating by a clinical staff member or physician about, for example, how the addict was doing that day. Physiologic measures or responses—for example, urine screens, hair analysis, naloxone challenge scores—should be obtained along with behavioral measures such as retention rates. Common or similar results across different domains sampled strengthen an NDA.

FDA's Pilot Drug Evaluation Division would permit four primary variables without penalizing for multiplicity. An approval may become difficult if effect is shown for only one variable in one population in one study only. The results obtained by analyzing a data set validated by FDA's Division of Scientific Investigations using a specific method (of analysis) are cross-validated by analyzing data using some other techniques to see whether the findings are robust. Implicit assumptions built into data collection, reduction, and analysis are evaluated. Knowledge of what took place at each step along the way—from preclinical work to analysis of phase III trials—is helpful. Trial designs that not only meet the requirements of a particular analytical technique to be used but also are robust toward dropouts, violation of protocol assumptions, and alternative analytical techniques are preferable. This is so because trials designed to prove efficacy may also be looked at to try to determine the dose, to evaluate adverse reactions, or to develop specific instructions for use for subpopulations. It is also important to look at what information may have been thrown away and what information may be so confounded that dose, duration of treatment, patient acceptability, specific adverse events, and management of patient dropouts are so distorted that the trial cannot be used to make a regulatory   decision.

Dr. Stein believed it important to evaluate the social impact of the proposed drug in these populations. How healthy and how productive the patients may be after the treatment is probably a primary variable for these populations. The endpoints should be reliable and quantifiable. Simple surrogate measures such as how frequently the drug is abused, what is the abuse pattern, and how much and what kind of drug is being abused are important. An acceptable analysis should be able to identify how each patient did during the treatment and what his or her contribution is to the overall analysis.

Dr. Gorodetzky commented about the use of four primary variables. The number of primary variables to be used will depend on the kind of experiment designed and whether it is aimed at the consumer, at the science, or at

medicine. Some kind of compromise is possible. A clinical trial is an experiment in which one has to think very specifically about the objectives and the operational manner in which one is going to attempt to reach those objectives. One may not want to do certain things in a given situation that might be interesting to do in another context. It is not as simple as choosing one variable or four variables; the question is how some very practical questions can be answered and how specific objectives can be drawn up for clinical trials.

The end product of an approved drug is a package insert aimed at the users-the practicing physicians and other scientists. The package insert communicates what should be expected from the approved drug. As Dr. Wright put it, what should be communicated to these users is fairly basic practical data: For example, is the patient going to be arrested less often? Is the patient going to be using drugs less? Is the patient going to come back to the clinic? If a package insert communicates information that is too complex, it would not be understandable to the users of a package insert.

As Dr. Wright pointed out, combination variables are good at supporting a fairly robust statistical outcome, but they can make it extremely difficult to go back to the original data for dose selection, to develop instructions for use for subpopulations, and to establish relationships between adverse events and treatment drugs.

There was some discussion about the retention rates in these trials. How should this variable be used? What does this variable mean? Dr. Vocci wanted to use this variable as an outcome measure not only because it is important for the analysis, but also because, if a treatment works for only a subpopulation, there is an interest in knowing the characteristics of that subpopulation. This variable might tell who is going to be a possible treatment success. Retention is important because, before patients can benefit from the treatment and, thus, start changing their behavior (other than drug-taking behavior), they must stay in the treatment for a certain length of time. This reflects on the effectiveness of a treatment program vs. the effectiveness of a drug. According to Dr. Gorodetzky, retention is a complex variable and may have more practical consequences than some of the other outcome variables.

Because treatment milieu differs substantially from one clinic to another, the largest treatment by investigator (clinic) interaction is likely to be discovered for retention in multicenter trials. People may drop out of these trials for different reasons: because of a 4-hour questionnaire they are asked to complete on the last day of the treatment; because the treatment failed for them; or because of how they get paid, how much they are paid, and when. Dropouts modify treatment effects in these trials in unknown ways.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Design of Clinical Trials for Treatment of Opiate Dependence: What Is Missing?

Ram B. Jain

INTRODUCTION

A typical trial to evaluate the safety and efficacy of a new pharmacotherapy for the treatment of drug dependence, including opiate dependence, would be double blind and would use one or more doses of the new pharmacological agent as well as a placebo and/or an active control as the alternative treatment arm. The primary outcome variable of interest will be the frequency and/or the amount of the addicting/abused opiates used by the subjects in the trial in different treatment arms. The only practical way to determine either the frequency or the amount of the addicting/abused opiates used by the addicts would be through self-reports. However, these self-reports are not likely to be very reliable. Consequently, the addicts are asked to provide urine samples as specified in the protocol. These urine samples are assayed to determine the presence and/or the amount of the addicting/abused opiates.

T_1 and T_2 are the two consecutive time points (figure 1) at which a subject provides urine samples for testing. If episodes A, B, and C are three independent episodes of opiate abuse, then A will not be detected at either T_1 or T_2, B will be detected at T, only, and C will be detected at both T, and T_2 since the amount of opiate abused at these episodes was different, and as such the duration for which opiates stay in the urine will be different. To detect episode A or to avoid underestimation of the frequency of opiate abuse, the urine samples should have been collected and assayed earlier; in other words, to avoid underestimation, the urine samples should be collected as *frequently as possible.* To avoid episode C being detected twice or to avoid overestimation of the frequency of opiate abuse, the urine samples should be collected *as infrequently as possible.* The phenomenon of two or more consecutive samples detecting the same episode of opiate abuse is called the *carryover* from one positive sample to another positive sample. There are substantial variations in drug-seeking behavior from one addict to another:
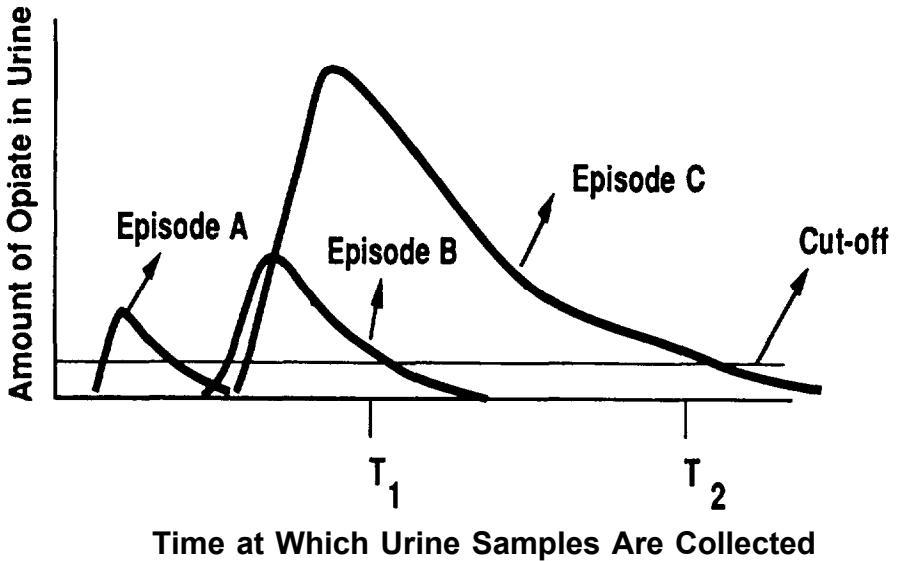
FIGURE 1. *Detection of drug abuse by urine assays*

Some abuse large amounts in relatively few episodes; some use small amounts in relatively large numbers of episodes; some abuse drugs during weekends only; and some use them every day. For this reason, it is difficult to determine whether two or more consecutive positive urine samples represent one or more episodes of drug abuse or, in other words, whether there is a carryover. Also, since the estimation of carryover is difficult, carryover or overestimation rather than underestimation of the frequency of opiate abuse is more of a concern. However, complete elimination of the probability of carryover may not be achievable. Hence, it is probably best to design the trials so that the probability of carryover from one positive urine to another positive urine is *minimized* and the probability of detecting an episode of drug abuse is *maximized.* This chapter provides suggestions as to how a trial can be designed to achieve this and what may still be missing.

The issues that reflect on the design of these trials can be studied under the following titles:

1. Sampling schemes used to obtain urine samples
2. Frequency and timing of the collection of urine samples
3. Qualitative vs. quantitative analysis of urine samples

SAMPLING SCHEMES USED TO OBTAIN URINE SAMPLES

In one of the earlier trials conducted to evaluate the safety and efficacy of LAAM, Ling and colleagues (1976) collected urine samples once a week using a *random time sampling scheme.* In a *random time sampling scheme,* although the subjects know how many times during a given week they will be asked to provide their urine samples, they do not know on which days of the week they will be asked to provide a urine sample. It is *randomly* decided who will provide a urine sample on which day of the week. For example, if the protocol calls for collection of one urine sample per week from each subject and if the urine samples are to be collected Monday through Friday only, 20 percent of the total subjects in the study will provide urine samples on Monday, 20 percent of the total subjects in the study (from the remaining 80 percent of the total subjects) will provide urine samples on Tuesday, and so on until all the subjects who have not provided their urine samples by Thursday will be asked to provide their urine samples on Friday. Consequently, the probability of a subject providing a urine sample will vary from day to day, ranging from zero to one. Consequently, this type of sampling scheme is not truly random. In addition, a subject X may provide a urine sample on Monday of one week and on Friday of the next week, thus being allowed free drug-seeking behavior for 10 days. On the other hand, a subject Y may provide a urine sample on Friday of one week and on Monday of the next week, thus being allowed *free* drug-seeking behavior for only 2 days. Thus, a *random time sampling* scheme has the potential to make alternate treatment groups incomparable for analysis. As said earlier, this sampling scheme is not truly random, but for lack of better terminology, it is called a *random time sampling scheme.* This type of sampling scheme was earlier advocated by Goldstein and Brown (1970).

Certain other types of *random time sampling schemes* are discussed in Harford and Kleber (1978) and Goldstein and Brown (1970). However, since these schemes are not in practical use, they will not be discussed further.

According to a report published by the Council on Scientific Affairs (1987), opiates stay in the urine for about 48 hours. Hence, unless urine samples are collected at less than 48-hour intervals, carryover is not likely to be a problem. Consequently, once-a-week, 5-days-a-week *random time sampling* is not likely to lead to carryover, but since an addict may be tested as far apart as 10 days, it certainly will lead to underestimation of the frequency of opiate abuse. But for twice and thrice a week, 5-days-a-week *random time sampling,* as can be seen from tables 1 and 2, the probability of being tested less than 48 hours apart, that is, on consecutive days, is 54.9 and 45.8 percent, respectively, which is likely to lead to a serious carryover. The probability of being tested more than 48 hours apart, that is, probability of underestimation, is 18.9 and 13.7 percent, respectively.

TABLE 1. *Probabilities of being tested in a twice-a-week, 5-days-a-week random time testing**

Probability of Being Tested on						Number of *Free* Drug-Seeking Days During the Week	Minimum (Maximum) Number of *Free* Drug-Seeking Days During 2 Weeks
M	T	W	T	F			
X	X				.1600000	5	5 (8)
X		X			.1142857	4	4 (7)
X			X		.0628571	3	3 (6)
X				X	.0628571	2	2 (5)
	X	X			.1142859	4	4 (7)
	X		X		.0628571	3	3 (6)
	X			X	.0628571	2	2 (5)
		X	X		.0857142	3	3 (6)
		X		X	.0857142	2	2 (5)
			X	X	.1885714	2	2 (5)

*Total probability of being tested on consecutive days=.5485713; probability of being tested more than 48 hours apart during the same week=.1885713.

Hence, *random time sampling* could render treatment groups incomparable for analysis and may result in serious underestimation of the frequency of opiate abuse and/or a serious carryover from one positive sample to another positive sample depending on the frequency of sampling.

To further dwell on the merits and demerits of *random time sampling,* another type of sampling scheme called *fixed time sampling* needs to be defined.

In a *fixed time sampling scheme,* all subjects are asked to provide urine samples on the same days of the week. In a double-blind, double-dummy clinical trial to compare the efficacy and safety of 8-mg sublingual doses of buprenorphine with 20- and 60-mg doses of methadone conducted at the Addiction Research Center of the National Institute on Drug Abuse (the ARC 090 trial), between September 1988 and May 1990, a *fixed time sampling scheme* was used to obtain urine samples three times a week on Mondays, Wednesdays, and Fridays. Because the urine samples were obtained at least 48 hours apart, the probability of carryover is minimal. According to Dr. Edward J. Cone (personal communication, July 1991) of the Addiction Research Center, the mean time to detect (cutoff=300 ng/mL) intramuscular administration of 6 mg of morphine by an enzyme-multiplied immunoassay technique (EMIT)

TABLE 2. *Probabilities of being tested in a three-times-a-week, 5-days-a-week random time testing**

M	T	W	T	F	Probability of Being Tested on	Number of *Free* Drug-Seeking Days During the Week	Number of *Free* Drug-Seeking Days During 2 Weeks
X	X	X			.1885714	4	4 (6)
X	X		X		.1487258	3	3 (5)
X	X			X	.0227026	2	2 (4)
X		X	X		.1090656	3	3 (5)
X		X		X	.0151350	2	2 (4)
X			X	X	.1142857	2	2 (4)
	X	X	X		.1090656	3	3 (5)
	X	X		X	.0151350	2	2 (4)
	X		X	X	.1142857	2	2 (4)
		X	X	X	.1600000	2	2 (4)

*Total probability of being tested on consecutive days=.457637; probability of being tested more than 48 hours apart during the same week=.1369883.

assay was 21.82 hours (n=5, SD=5.34). Given that the urine half-life of morphine is 4 to 6 hours, on the average, up to 96 mg of morphine can be consumed by an addict during one episode and still result in only one positive urine if the consecutive urines are collected and assayed at least 48 hours apart. However, since Friday and Monday samples were collected 72 hours apart, the potential for underestimation is certainly there, but this is likely to happen only when opiates are abused on Fridays but not on Saturdays and Sundays. At worst, the addicts have 3 *free* days of drug-seeking behavior. But because everybody has the same number of *free* days uniformly across the whole study period, the comparability of different treatment groups is maintained.

The strongest argument in favor of *random time sampling* is that the addicts try to avoid drug abuse detection, and as such, if they know they will be tested, they will not show up for their scheduled visits. In certain special treatment situations in which a positive result is associated with certain contingencies, this might be true, but in a clinical trial environment there is no reason to expect any such contingencies. As such, the argument to use *random time sampling* is merely philosophical, with no advantage and many drawbacks, including a substantial potential to render the data nonanalyzable. If a protocol calls for administrative withdrawal after a certain number of positive urines, the addict may be switched to an alternate, possibly more beneficial treatment

rather than being withdrawn, and makeup urines may be collected on days following a missed visit; these makeup urines may or may not be used in the analysis, In addition, there are no published data to suggest that such a practice does occur in a noncontingent treatment environment.

Hence, a *fixed time sampling* should be the design of choice.

FREQUENCY AND TIMING OF THE COLLECTION OF URINE SAMPLES

When and how frequently the urine samples should be collected depends on the kinetics of the drug of abuse and the sensitivity of the assay used to analyze the urine samples. For heroin, with a cutoff of 300 ng/mL, a sample every 48 hours seems to be the optimal choice, because as pointed out by the Council on Scientific Affairs (1987), heroin stays in the urine for about 48 hours provided EMIT-type assays are used. This is likely to *minimize* the probability of carryover and *maximize* the probability of detecting an episode of opiate abuse.

With a lower cutoff and/or a more sensitive assay such as gas chromatography/ mass spectrometry, the samples *may* have to be collected and assayed more infrequently. Otherwise, the probability of carryover may be increased. However, this may decrease the probability of detecting an episode of drug abuse.

Also, for shorter acting drugs, the samples may have to be collected more frequently. For longer acting drugs, they may have to be collected more infrequently.

The timing of sample collection should be such that the days of heavy use do not go undetected. For example, to detect use on weekends, it may be necessary to collect the first sample of the week on Monday.

In summary, the decision of when and how frequently the samples should be collected should be made by a joint team: a statistician, who should ensure that the probability of carryover is minimized and the probability of detecting the drug abuse is maximized to the degree possible; a pharmacologist/ pharmacokineticist, who should ensure that reliable information on the kinetics of the drug of abuse is available and is provided to the statistician; and a physician/clinician, who is adequately informed of the pattern of drug abuse and should be primarily responsible for the timing of sample collection.

QUALITATIVE VS. QUALITATIVE ANALYSIS OF URINE SAMPLES

Currently, the clinical trials in the drug abuse area are designed to estimate the frequency of drug abuse and not the amount of drug abuse. However, a replacement drug may decrease the frequency of drug abuse, but the addicts may still be using the same amount of the drug (of abuse), though in a smaller number of episodes. The amount of drug abuse may be estimated by analyzing the urine samples quantitatively rather than qualitatively, that is, by estimating the amount of the drug of abuse in the urine, rather than just the presence or absence of the drug of abuse. However, a real-life relationship between the amount of drug present in the urine and the actual amount of drug consumed is confounded by many factors.

A relationship between the amount of drug present in the urine and the actual amount of drug consumed may be established in laboratory experiments, and an inference can be drawn about the amount of drug consumed from the amount of drug present in the urine. However, a relationship established in the laboratory is not likely to hold in real-life situations because of the uncertainty of the timing of the episodes of drug abuse, the variations in the purity of drugs of abuse with different geographic locations and times, the effect of multiple episodes of drug abuse on the metabolism of these drugs, the interactions between multiple drugs of abuse consumed by the addicts in same or different episodes, the differences in frequency and timing of drugs abused by the addicts, and so on. And, of course, how accurately this relationship can be determined will also depend on the accuracy of the quantitative assays used to analyze urine samples. In addition, instead of urine samples, plasma samples may be better determinants of this relationship, but once again, this relationship too will be confounded by the same factors that confound this relationship for urine samples.

At best, a relationship between the amount of drug present in the urine or plasma samples and the actual amount of drug abused is very complex and not easy to capture in real-life situations. However, a joint effort by statisticians, pharmacokineticists, and physicians/clinicians to model this relationship is likely to be fruitful.

It must also be mentioned that the estimation of the amount of drug abuse should not be done in lieu of the estimation of the frequency of drug abuse. Both should be done simultaneously. Because of the strong relationship between the frequency of intravenous use and human immunodeficiency virus infection, it is of paramount importance that the replacement drugs should decrease the frequency as well as the amount of drug abuse.

WHAT IS MISSING?

1. The statistical/pharmacokinetic methods/design to model the relationship between the amount of drugs present in the urine or plasma samples and the actual amount of drugs abused is missing.

2. The present methods to estimate the frequency of drug abuse provide, at best, a lower bound on the frequency of drug abuse because of:

   • The inability to detect possible multiple episodes of drug abuse during the time two consecutive urine samples are collected, and

   • The need to do *infrequent sampling* to *minimize* the carryover from one positive sample to another positive sample.

3. The probability of carryover is not entirely eliminated, and the degree of carryover is not known. It will be helpful if methods/techniques can be developed to ascertain whether multiple, consecutive positive samples are due to one or multiple episodes of drug abuse. This *may,* for example, be done by using self-reported episodes of drug abuse during the time consecutive urine samples are collected.

REFERENCES

Council on Scientific Affairs, Scientific issues in drug testing. *JAMA* 257:3110-3114, 1987.

Goldstein, A., and Brown, B.W. Urine testing schedules in methadone maintenance treatment of heroin addiction. *JAMA* 214:311-315, 1970.

Harford, R.J., and Kleber, H.D. Comparative validity of random-interval and fixed-interval urinalysis schedules. *Arch Gen Psychiatry* 35:356-359, 1978.

Ling, W.; Charuvastra, V.; Kaim, S.C.; and Klett, C.J. Methadyl acetate and methadone as maintenance treatments for heroin addicts. *Arch Gen Psychiatry* 33:709-720, 1976.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Comments on "Design of Clinical Trials for Treatment of Opiate Dependence: What is Missing?" by Jain

Sudhir C. Gupta

This chapter discusses the following three important issues in the design of clinical trials for opiate dependence:

1. Random vs. fixed time sampling scheme for collecting urine samples
2. Frequency and timing for collecting urine samples
3. Estimating the amount of drug abuse in addition to the frequency of drug abuse

SAMPLING SCHEME FOR COLLECTING URINE SAMPLES

As discussed by Dr. Jain, the main problem with using a random time sampling scheme is that the methods for analyzing the data obtained using this scheme may not be available. This means that suitable methods should first be developed before the analysis of the data can be carried out. As pointed out by Dr. Jain, this approach is not recommended. The trial should be designed so as to allow an efficient interpretation of the data. A fixed time censoring scheme is thus recommended.

The strongest argument in favor of random time sampling is that addicts try to avoid drug abuse detection. In a fixed time sampling scheme if they know that they will test positive because of drug abuse, they may not show up for their scheduled visits. However, Dr. Jain has pointed out that this is not to be expected in this trial because subjects who are known to be drug addicts do not have anything to gain by avoiding detection of drug abuse. In a fixed time sampling scheme all the subjects are required to provide urine samples on each of the scheduled days. Sometimes it may become necessary to use a random time sampling scheme if enough resources are not available to handle all the subjects in one day. If a random time sampling scheme is to be used under such circumstances, then it should be modified to yield truly random samples as indicated below.

Suppose the protocol calls for collection of two urine samples per week from each subject. Then there should be an equal probability for a subject to be tested on any 2 of the 5 days of the week. Let MTh denote that a subject is to be tested on Monday and Thursday, etc. A subject may be tested on MT, MW, MTh, MF, TW, TTh, TF, WTh, WF, or ThF, resulting in 10 possibilities as pointed out by Dr. Jain. A subject should be assigned to one of these 10 possibilities randomly. This random assignment should be done separately for each week, and it should not be known to the subjects in advance of their urine collection.

In the case of two urine samples per week, the expected number of free drug-seeking days is 3.15 using table 1 of Dr. Jain's chapter. For the above suggestion the probability is 0.10 for a subject to be tested on any of the 10 possible pairs of days. The expected number of free drug-seeking days is then 3.0.

A similar method can be used for three urine samples per week, reducing the expected number of free drug-seeking days to 2.2. The corresponding expected number is 2.74 from Dr. Jain's chapter.

FREQUENCY AND TIMING FOR COLLECTING URINE SAMPLES

As pointed out by Dr. Jain, the frequency of collecting urine samples should be determined so as to minimize the probability of carryover and to maximize the probability of detecting opiate abuse. As discussed in Gupta (1991), a model that incorporates subject and carryover effects can be developed using the approach of Bonney (1987). However, in this approach the subject effects and carryover effects are confounded, and a separate estimate of carryover effect is not provided. This does not seem to be a serious limitation.

ESTIMATING THE AMOUNT OF DRUG ABUSE IN ADDITION TO THE FREQUENCY OF DRUGABUSE

Dr. Jain has clearly discussed the problems associated with estimating the amount of drug abuse in addition to the frequency of abuse. As pointed out by Dr. Jain, currently the clinical trials in this area are designed to estimate the frequency of drug abuse and not the amount of drug abuse. If the addict tests positive for drug abuse, then it is important to find out the extent to which the drug was abused. In other words, it is important to know if a replacement therapy is effective in reducing the total amount of drug abused in addition to reducing the frequency of drug abuse. A relationship between the amount of drugs present in the urine and the amount of drug consumed by the addict may be established in laboratory experiments, from which an estimate of the amount of drug consumed may be obtained. However, as Dr. Jain has clearly pointed out, such estimates are confounded by many factors. Therefore, such

estimates derived using the results obtained in the laboratory will not be precise. Under these circumstances it will be best to study the extent rather than the exact amount of drug abuse. Let us assume, for example, that the extent of drug abuse is categorized as low, medium, or high.

Let the outcome variable Y be coded as 0 if the assay shows absence of abused opiates in the urine. Similarly, Y = 1, 2, 3 will be used to denote that the assay shows the extent of drug abused to be low, medium, and high, respectively. Since the outcome variable takes more than two distinct values, an appropriate polytomous logistic regression model can be developed for comparing the probabilities under different treatments after adjusting for the effects of covariates. A patient provides repeated observations up to a maximum of 17 weeks. Since each dose of a treatment drug provides one observation, a maximum of 51 replications for a treatment can be obtained for any patient. These observations from the same patient will not be independent. Thus, conditional probabilities will be used under the polytomous logistic regression setup.

Suppose that there are n_i observations for the ith patient, which are denoted by Y_{ij}, and let $X_{ij} = (X_{1ij}, X_{2ij}, \ldots, X_{p\,i\,j})'$ denote the vector of covariates associated with Y_{ij}, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n_i$. Let

$$Y_{ij}^* = [Y_{i1}, Y_{i2}, \ldots, Y_{i,j-1}]'$$

$$y_{ij}^* = [y_{i1}, y_{i2}, \ldots, y_{i,j-1}]'$$

$$P(Y_{i1} = y_{i1} | X = X_1) = \pi_{1y_{ij}}$$

$$P(Y_{ij} = y_{ij} | Y_{ij}^* = y_{ij}^*, X = X_i) = \pi_{jy_{ij}}$$

$$y_{ij} = 0, 1, 2, 3$$

$$i = 1, 2, \ldots, m$$

$$j = 1, 2, \ldots, n_i$$

Following the approach of Bonney (1987) as discussed by Gupta (1991) for the case of dichotomous outcome variables, the conditional probabilities as defined above can be modeled by considering $\acute{Y}_{ij}^*$ as covariates. Let

$$\beta_\ell = (\beta_{1\ell}, \beta_{2\ell}, \ldots, \beta_{p\ell})'$$

$$\eta_{j\ell}^* = (\eta_{1\ell}, \eta_{2\ell}, \ldots, \eta_{j-1,\ell})'$$

$$\pi_{10} = \frac{1}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta_\ell' X_{i1}}\right)}$$

$$\pi_{j0} = \frac{1}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta'_\ell X_{ij} + \eta^{*\prime}_{ji} Y^*_{ij}} \right)}$$

$$\pi_{1s} = \frac{e^{\alpha_s + \beta'_s X_{i1}}}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta'_\ell X_{i1}} \right)}, \quad s = 1, 2, 3$$

$$\pi_{js} = \frac{e^{\alpha_s + \beta'_s X_{ij} + \eta^{*\prime}_{js} Y^*_{ij}}}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta'_\ell X_{ij} + \eta^{*\prime}_{ji} Y^*_{ij}} \right)}, \quad s = 1, 2, 3$$

$$i = 1, 2, \dots, m$$
$$j = 1, 2, \dots, n_i$$

The logits for comparison with $Y = 0$ are thus obtained as given below. The logits for comparing with other values of Y can be written down in a similar way.

$$\log \left(\frac{\pi_{1s}}{\pi_{10}} \right) = \alpha_s + \beta'_s X_{i1}$$

$$\log \left(\frac{\pi_{js}}{\pi_{j0}} \right) = \alpha_s + \beta'_s X_{ij} + \eta^{*\prime}_{js} Y^*_{ij}$$

$$s = 1, 2, 3, \ i = 1, 2, \dots, m, \ j = 1, 2, \dots, n_i.$$

Following Gupta (1991), finally the model can be written as given below.

$$\pi_{j0} = \frac{1}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta'_\ell X_{ij} + \gamma'_\ell W_{ij} + \delta'_\ell Z_{ij}} \right)}$$

$$\pi_{js} = \frac{e^{\alpha_s + \beta'_s X_{ij} + \gamma'_s W_{ij} + \delta'_s Z_{ij}}}{1 + \sum_{\ell=1}^{3} \left(e^{\alpha_\ell + \beta'_\ell X_{ij} + \gamma'_\ell W_{ij} + \delta'_\ell Z_{ij}} \right)}$$

where

$$W_{ij} = (W_{1ij}, W_{2ij}, \dots, W_{n-1,ij})'$$

$$Z_{ij} = ((Z_{1ij}, Z_{2ij}, \dots, Z_{n-1,ij})'$$

$$\gamma_\ell = (\gamma_{1\ell}, \gamma_{2\ell}, \dots, \gamma_{n-1,\ell})'$$

$$\delta_\ell = (\delta_{1\ell}, \delta_{2\ell}, \dots, \delta_{n-1,\ell})'$$

$$W_{uij} = \begin{cases} -1 & \text{if } Y_{iu} = 0 \text{ and } u < j \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{uij} = \begin{cases} 1 & \text{if } Y_{iu} = 1 \text{ and } u < j \\ 0 & \text{otherwise} \end{cases}$$

$$i = 1, 2, \dots, m, \ j = 1, 2, \dots, n, \ u = 1, 2, \dots, n-1,$$

and n denotes the maximum number of repeated observations possible for a subject, in the present case $n = 51$.

The parameters of the model are estimated using the method of maximum likelihood. Note that in practice certain interactions may be needed to be included in the model.

REFERENCES

Bonney, G.E. Logistic regression for dependent binary observations. *Biometrics* 43:951-973, 1987.

Gupta, S. "Analysis of Treatment for Drug Dependence Data Using Logistic Regression," Paper presented at the Technical Review Committee Meeting on Statistical Issues in Clinical Trials for Treatment of Drug Dependence, National institute on Drug Abuse, December 2-3, 1991.

AUTHOR

Sudhir C. Gupta, Ph.D.
Associate Professor
Division of Statistics
Northern Illinois University
DeKalb, IL 60115-2888

Rejoinder

Ram B. Jain

In his comments, Dr. Gupta seems to suggest that if subjects can be tested on any *x* (2) days of the week in an *x* (2)-days-a-week *random time sampling* with the same probabilities, this sampling scheme can be considered to be *truly* random. I am not quite sure if this is essentially true. Once the subject has been tested on any 2 days during a given week, he or she will have zero probability of being tested again and will be free to abuse the drugs. If the sampling must be *truly* random, he or she should have equal probability of being tested on any given day of the week on which urine samples are scheduled to be collected.

In addition, the biggest problem with random time sampling schemes is not that they are not truly random. Their biggest problem is that in these kinds of sampling schemes some subjects are tested too soon after their previous test and some are tested too long after their previous test. This leads to the problem of carryover and the loss in ability to detect an episode of drug abuse.

Dr. Gupta also suggested in his written comments and during the meeting that carryover can be incorporated in the model based on a previous positive urine. Unfortunately, two or more consecutive positive tests do not always indicate a carryover. An assumption that two or more consecutive positive urines always indicate a carryover is likely to lead to underestimation of the probability of drug abuse. However, subjects' self-reports about the past drug abuse may be used to make a decision as to whether two or more consecutive positive urines indicate a carryover.

Dr. Gupta indicates in his comments that the logistic regression model he proposed can be used to incorporate carryover effect, but subject effect and carryover effect will be confounded and a separate estimate of carryover effect will not be available. I believe this is a serious limitation. Dr. Gupta does not agree.

In the absence of methodology to exactly estimate the amount of drug abuse, Dr. Gupta suggests a logistic regression model to estimate the extent of drug abuse categorized as low, medium, or high. This certainly would be an idea

worth pursuing. However, to depend on the assays to categorize as low, medium, or high drug use will be somewhat unreliable because the amount of drug present in the urine at the time of testing not only depends on the amount of drug consumed but also on the time the drug was consumed since the last urine sample was collected. The timing of the episodes of drug abuse since the last urine sample was collected is not likely to be known with much accuracy.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical   Statistician
Biometrics Branch
Medications   Development   Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Summary of Discussion: "Design of Clinical Trials for Treatment of Opiate Dependence: What Is Missing?"

Ram B. Jain

Dr. Murphy expressed the concern that by asking the subjects to come to the clinic to provide urine specimens three times a week in a *fixed time sampling,* we are creating a compliance problem. Even if there are no negative contingencies associated with the positive urines, certain subjects will not enroll because they will have to provide urine specimens three times a week. Retention rates might be adversely affected in a three-times-a-week *fixed time sampling.* I pointed out that if the subjects are asked to come to the clinic three times a week, the problem will be the same whether it is a *fixed time* or *random time* sampling. If Dr. Murphy's suggestion was to ask them to come to the clinic once a week, we would be knowingly collecting fewer data than are needed to estimate treatment effects. A lot of episodes will go undetected in once-a-week sampling. This will defeat the purpose of the clinical trials, which is to estimate the treatment effect from several different treatments and compare them for efficacy.

As pointed out by Dr. Vocci, the primary purpose of a clinical trial is to evaluate the pharmacological effect of the treatment in a natural setting rather than to manipulate dropout rates and/or drug abuse rates by introducing negative contingencies into the trial, and there were none in the ARC 090 trial. In addition, as Dr. Johnson pointed out, the subjects were asked to hold medication under the tongue for 10 minutes in this trial; this alone would create some compliance problems. Since the subjects come to the clinic every day for their medication and other procedures, asking them to provide urine specimens for 3 of these 7 days should create no additional compliance problems.

Dr. Jack C. Lee expressed concern about differences (periodicity) in missed visit rates, percent positive rates, etc., on the different days (Monday vs. Wednesday vs. Friday) the urine specimens were obtained, as was seen in Follmann and colleagues' chapter. Dr. Lee suggested that, in place of

collecting the samples a fixed number of times on fixed days of the week, a *random* scheme may be adopted so that the expected number of tests during a week may be fixed (e.g., three), but the samples may be collected on different days of the week and a different number of times during different weeks. This may help smooth out the periodic (cyclic) effect seen in the data. I pointed out that by using such a scheme we will run into the problem of testing some subjects too soon and some too long after the previous test, and this may lead to carryover and/or avoiding detection of certain drug abuse episodes. In addition, differences in missed visit rates and/or treatment effect across the different days the urine specimens are collected provide some useful information, and such differences should be expected. Even if the treatment is working, subjects can be expected to abuse more during weekends than during weekdays because of social pressures, etc.; the effectiveness of treatment may be expected to diminish during weekends.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Efficacy of Urinalysis in Monitoring Heroin and Cocaine Abuse Patterns: Implications in Clinical Trials for Treatment of Drug Dependence

Edward J. Cone and Sandra L. Dickerson

INTRODUCTION

Human self-administration of drugs of abuse begins a series of biochemical and pharmacologic events that culminates in the alteration of an individual's mood state. The physical and chemical processes that ultimately determine the extent of drug effect, that is, how much active drug accumulates in the drug-receptor biophase, also serve to terminate the drug's actions. The primary processes responsible for the appearance and termination of these effects are absorption, distribution, metabolism, and excretion. The time courses of these processes are illustrated in the generic example shown in figure 1, panel A, for the appearance and disappearance of a drug in urine. Urine drug levels typically increase rapidly after administration and peak and decline at a slower rate. In this example, the analytical technique used to measure drug levels has an assigned cutoff of 300 ng/mL. Cutoffs are used to categorize urine specimens as positive or negative for drug; they are assigned based on analytical factors such as assay precision and reproducibility and on therapeutic considerations such as drug potency and rate of excretion. For opiates and cocaine, 300 ng/mL was selected as the screening cutoff for use in Urine testing of Federal employees (Mandatory guidelines 1988). This cutoff is in common use throughout Federal and private-sector employee testing programs and in treatment programs.

As shown in figure 1, urine drug levels had declined to the cutoff by 36 hours after drug administration. All urine specimens obtained prior to that time would have tested positive. This is an ideal example of a detection time for a drug obtained by urine testing; this time interval represents the time elapsed from drug administration to excretion of the last positive specimen. This concept is extremely useful when implementing a drug testing program for treatment of drug addicts or conducting a clinical trial for a new medication. In these

A. **Metabolism and Excretion**

Distribution

Absorption

Concentration (ng/mL) axis: 1,000, 800, 600, 400, 200, 0

300-ng/mL cutoff

Hours axis: 0, 20, 40, 60

Drug

Detection Time

B. **Detection Time**

Detection time =36 hours

Hours axis: 0, 20, 40, 60
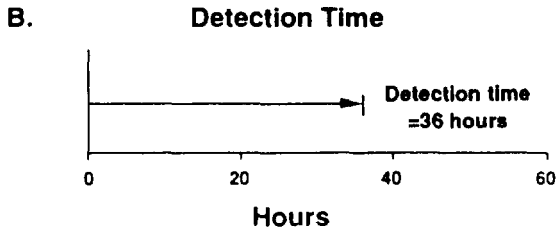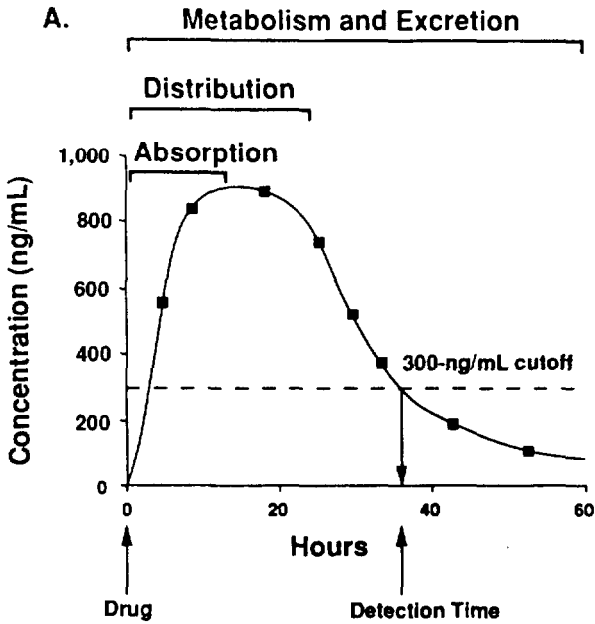
FIGURE 1. *Illustration of drug absorption, distribution, metabolism, and excretion phases for a short-acting drug during excretion in urine (panel A) and relationship of detection time to cutoff selection (panel 6)*

situations, it is vitally important to know whether illicit drugs are being used. Urine testing is recognized as the most objective means of diagnosing recent drug use (Hat-ford and Kleber 1978). Detection of a short-acting psychoactive substance such as heroin or cocaine in urine obviously indicates recent usage. In clinical trials that test drug-abusing subjects, the absence of drug use

47

generally indicates a successful outcome, whereas multiple drug use patterns indicate failure. In many situations, the degree of success may be judged on the basis of the number of positive urine test results obtained during the course of the clinical trial.

This chapter reviews the usefulness of detection times in relation to the conduct of clinical trials of new medications designed for the treatment of drug dependence. A fixed-interval urinalysis schedule is proposed, which optimizes the chances of detection of cocaine and/or heroin use while minimizing the risk of overlap of test results from a single episode of drug self-administration. Although random-interval schedules have been proposed as being more efficient for detection of illicit drug use in treatment (Goldstein and Brown 1970; Harford and Kleber 1978), it appears unlikely that the random-interval schedule would provide sufficient coverage for estimation of the extent of drug use. Hence, only fixed-interval schedules are considered in this chapter.

INFLUENCE OF DOSE AND CUTOFF ON DETECTION TIMES

There are many pharmacologic and chemical factors that influence detection times (Gorodetzky 1977). Pharmacologic factors include drug dose, route of administration, pH of the biological fluid, and individual differences in rates of metabolism and excretion. Chemical factors that relate to the analytical technique used for drug detection include selection of the cutoff, assay precision, specificity, and accuracy. The authors have systematically studied the influence of two of these factors, cutoff and dose, on detection times of cocaine and opiates. Figure 2 illustrates the influence of cutoff on the detection time of cocaine following administration of a 20-mg intravenous (IV) dose of cocaine hydrochloride. As the cutoff is lowered (greater assay sensitivity), the detection time increases (drug is detected longer). Figure 3 is a combined plot illustrating the changes in detection times of cocaine (panel A), morphine (panel B), heroin (panel C), and codeine (panel D) on a linear scale. The incremental changes in detection time with cutoff appear to be linear for cocaine and codeine and curvilinear for morphine and heroin. Regardless of the shape of the curve, these increases in detection time with the lowering of the cutoff are substantial. Clearly, the selection of the cutoff will have a major impact on the period of drug detectability. Consequently, outcome comparisons of clinical trial results between participating centers can be made only when identical cutoffs are utilized. In most cases, the recommended cutoffs by the U.S. Department of Health and Human Services Mandatory Guidelines (Mandatory guidelines 1988) should be used since most commercial assays are targeted toward and perform best at these concentrations. Also, since substantial differences occur in immunoassay specificity from different commercial vendors (Cone and
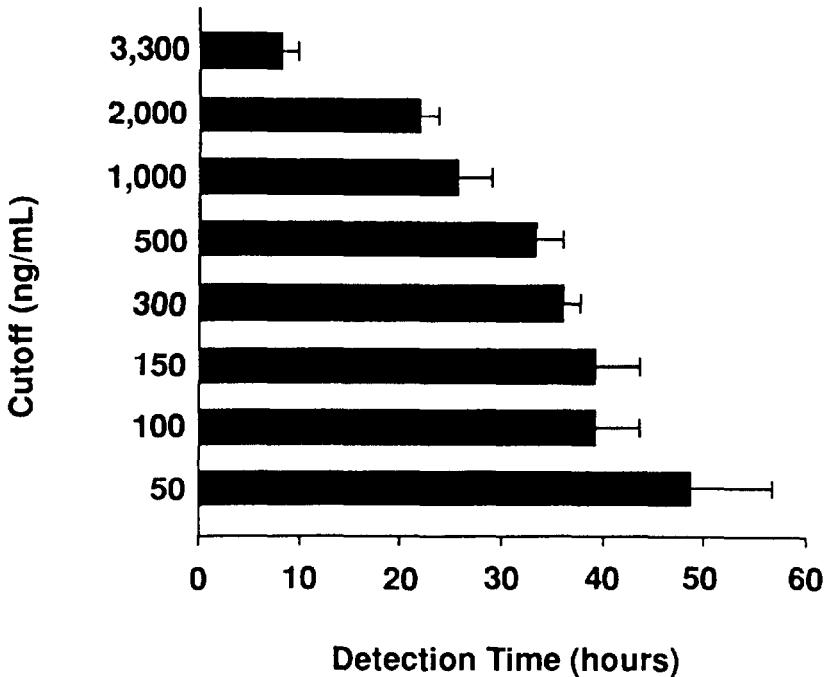
FIGURE 2. *Mean detection times at different cutoffs for cocaine (20-mg IV dose) by EMIT d.a.u. cocaine analysis. Error bars represent standard error of the mean (n=4).*

Mitchell 1989; Cone et al. 1992), identical urinalysis technology should be employed by each participating center.

For pharmacokinetic reasons, there is a log-linear relationship between drug dose and detection times. Consequently, detection times increase by one half-life each time the drug dose is doubled. For example, if the detection time of a 3-mg dose of heroin is 14.5 hours (300ng/mL cutoff) and the urinary excretion half-life of morphine (the analyte tested for heroin use) is approximately 6 hours, the detection time should increase to 20.5 hours when a person administers a 6-mg dose. The data illustrated in the bar graph in figure 4 indicate that the mean detection time of heroin for six subjects actually increased to 21.8 hours. For morphine, the mean detection time (n=6) increased from 34 hours for a 10-mg dose to 44 hours for a 20-mg dose. For codeine, the mean detection time (n=4) increased from 48 hours for a 60-mg dose to 54 hours for a 120-mg dose. These data are convincing evidence that a log-linear relationship exists
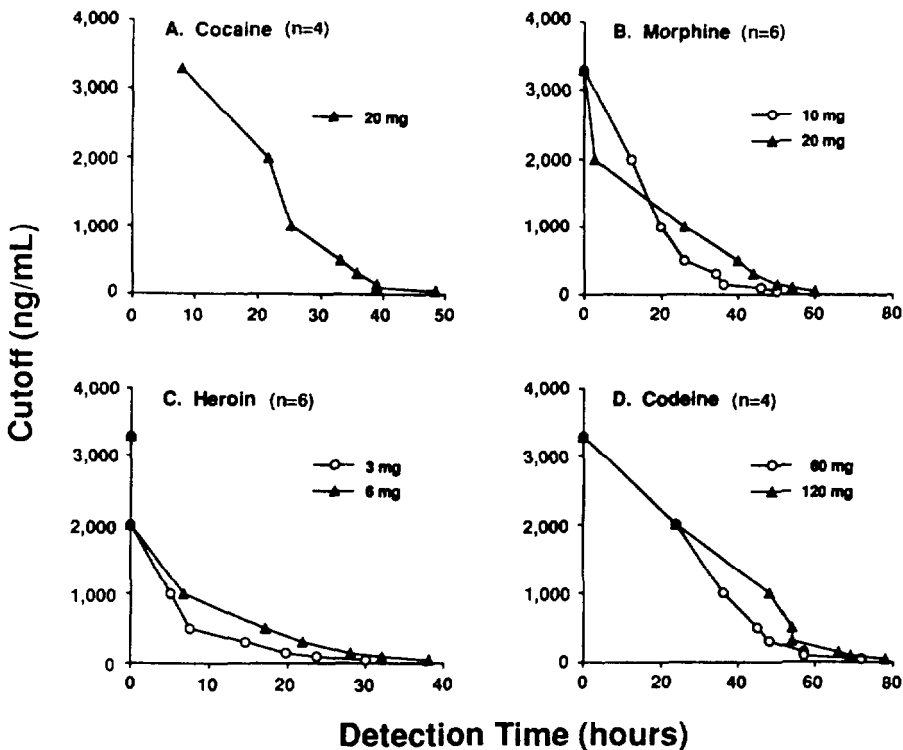
FIGURE 3. *Relationship of cutoff to detection times for cocaine, morphine, heroin, and codeine*

between dose and detection time for these drugs. Because of this relationship, *changes in drug dose by the user alters detectability of drugs by urinalysis only slightly.* This is fortuitous since the magnitude, frequency, and nature of illicit drug use by participating subjects are major variables in controlled clinical trials.

FREQUENCY OF TESTING VS. "SAFE TIME"

The dilemma in deciding how many times per week to test subjects arises from the need to maximize the chances of drug detection while minimizing the chances of counting a single drug use incident as two episodes and also minimizing the financial costs to the program and the inconvenience to subjects and staff. Figure 5 illustrates the amount of time during a week that a subject can use cocaine without being detected if testing is performed once per week. In this example, the mean detection time of 35.8 hours for cocaine (Cone et al.

50
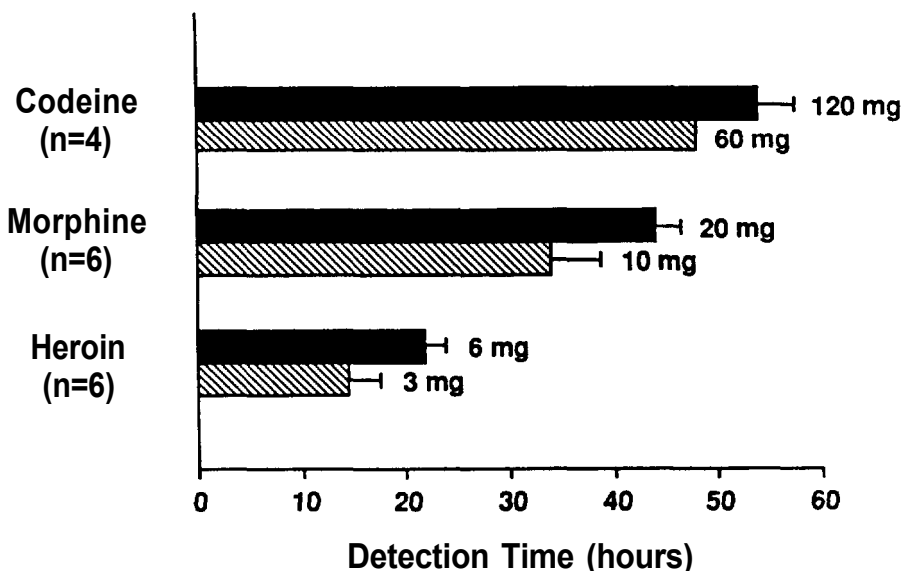
FIGURE 4. *Mean detection times by EMIT d.a.u. analysis vs. dose (300-ng/ mL cutoff) of codeine, morphine, and heroin*

1989) is used; hence, if the subject uses cocaine in this time period prior to testing on Friday, drug use will be detected. Drug use during any other part of the week will not be detected, resulting in a total of 132.2 hours of "safe time."

Obviously, the amount of safe time varies with the drug testing schedule. Figure 6 illustrates the amount of safe time arising from different weekly schedules of cocaine testing. If testing were performed 7 days a week, nearly all drug use would be detected; however, this is impractical in most cases because of subject, staff, and financial limitations. Furthermore, detection times for cocaine and heroin can extend beyond 24 hours; hence, drug excretion following a single use would extend through the next testing session, and a single use would be mistakenly counted twice. In contrast, an infrequent testing schedule would miss a substantial amount of illicit use and the urine data would be fallacious.

Figure 7 illustrates the amount of safe time (%week undetected) for cocaine and opiates for testing schedules varying from zero to 7 test days a week. It is apparent for three of the four drugs that there is an inflection in the graph at the 3-days-per-week schedule. Only heroin showed a linear decline. This

Detected Time

Safe Time   **Safe Time**

Detection Time=
Time Undetected=60.2 hours   35.8 hours   Time Undetected=72 hours

M   T   W   Th   F   Sa   S   M

Drug   Urine Test
Administration
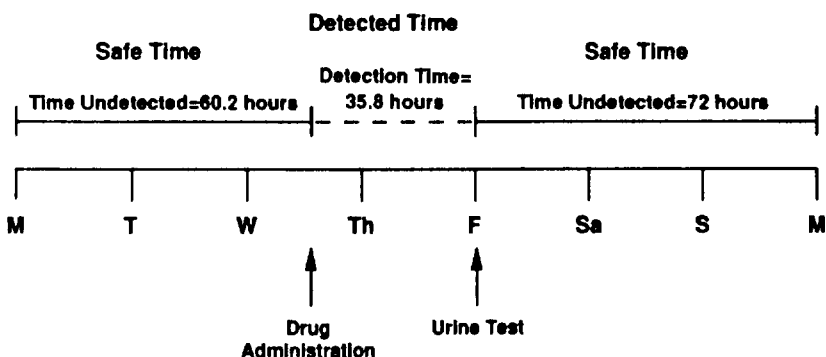
FIGURE 5. *Illustration of safe time and detected time for a once-a-week testing schedule for cocaine. It should be noted that, a/though testing was performed on Friday in this example, the results would have been identical for any other test day of the week.*

is likely due to the small doses employed in the heroin study resulting in minimal detection times. For morphine, codeine, and cocaine, the amount of safe time declined rapidly from zero to the 3-days-per-week testing schedule. Thereafter, the amount of safe time decreased more slowly. Consequently, it appears that a 3-days-per-week schedule provides the most parsimonious approach to testing when considering how to minimize both safe time and excretion overlap at the same time.

RANDOM DRUG USAGE VS. DIFFERENT URINE TESTING SCHEDULES

If a drug-abusing subject self-administers a single dose of cocaine during the course of a week, will the selected drug testing schedule detect drug abuse? This question was tested by generating four sets of 100 randomly selected times during a given week in which a subject might administer cocaine. No restrictions were placed on the time of drug use. A mean detection time of 35.8 hours was used in the calculation of *%drug episodes detected.* Individual and mean data are shown in table 1 for different testing schedules. The mean %drug episodes detected increased in a linear fashion from zero (no test days) to 63 percent with a 3-days-per-week schedule (Monday, Wednesday, Friday). Thereafter, the increase slowed and culminated in 100 percent of drug episodes detected with a 7-days-per-week schedule. Carryover from test to test as a result of the single drug dose did not begin to occur until the number of test days increased to 4 days per week. Thereafter, carryover increased substantially to nearly 50 percent with a 7-days-per-week schedule.
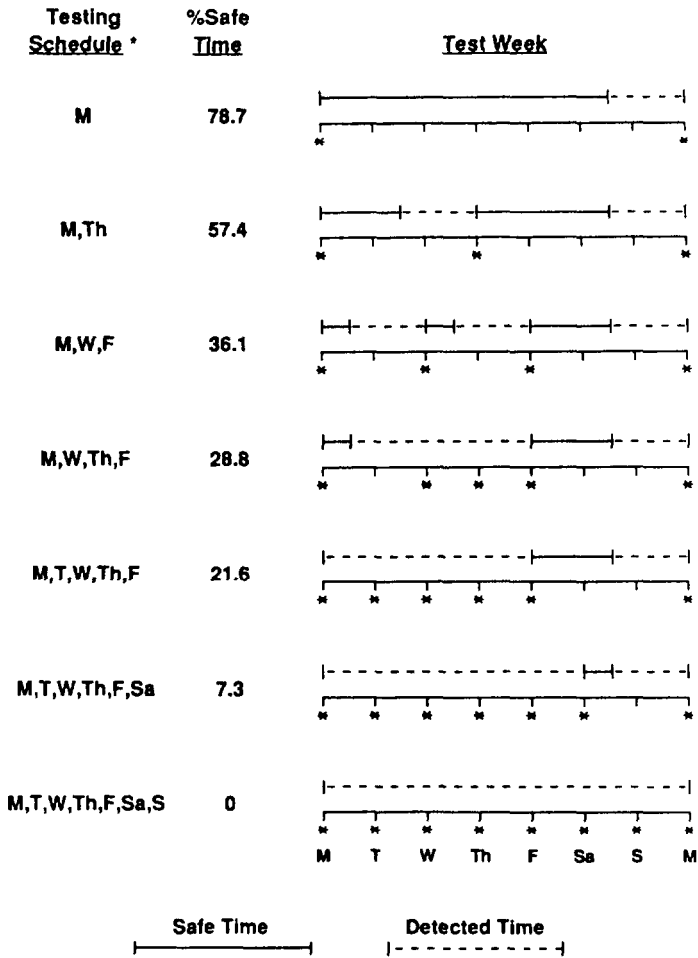
FIGURE 6. *The relationship of %safe time to the urinalysis testing schedule. Test days are indicated by an asterisk.*

A second analysis of urine testing schedules was performed by simulating two random cocaine uses occurring during the same week. The time between the two doses was varied from 6 hours to 84 hours. Sets of 100 randomly selected times, separated by the minimum interval between cocaine use, were generated. The effectiveness of testing three times per week was compared with testing only once per week. The numbers of times that two uses resulted

A. Cocaine (n=4)

▲ 30 mg

B. Morphine (n=6)

○ 10 mg
▲ 20 mg

C. Heroin (n=6)

○ 3 mg
▲ 6 mg

D. Codeine (n=4)

○ 60 mg
▲ 120 mg

%Week Undetected

Days/Week Tested
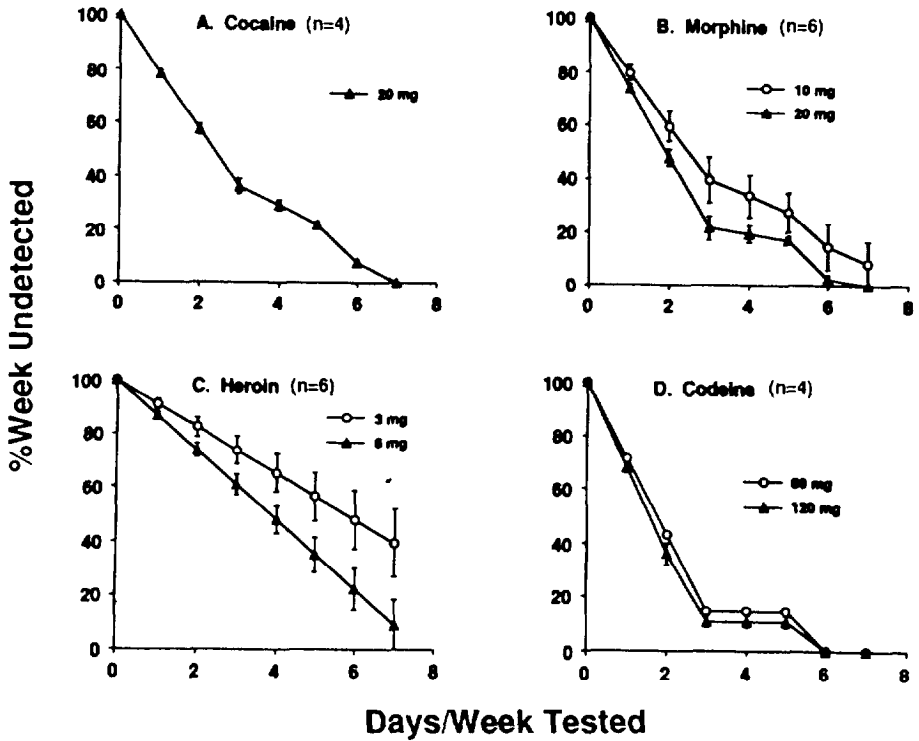
FIGURE 7. *Relationship of drug testing schedules to %week undetected for cocaine, morphine, heroin, and codeine (EMIT d.a.u. analysis, 300-ng/mL cutoff)*

in 0, 1, and 2 positive results are shown in table 2 along with the number of times that two uses occurred within the same detection time period resulting in a single positive result. When the testing schedule called for only 1-day-per-week testing, a substantial amount of drug use went undetected. The number of times that no drug use was detected varied from 64 to 39 percent depending on the time interval between uses. Positive results ranged from 36 to 61 percent. There were only a few occurrences of random multiple drug use occurring within the same detection time.

With a 3-days-per-week testing schedule (Monday, Wednesday, Friday), detection efficiency increased substantially over the 1 -day-per-week testing schedule. The number of times that no positive results were obtained by the 3-days-per-week schedule varied from 6 to 16 percent. Single positive results

TABLE 1. *Effect of urinalysis testing schedules on detection of a single cocaine use during a week of testing**

| Urinalysis Testing Schedule | Tests/ Week | %Drug Episodes Detected | | | | | Average Single Drug Use Episodes (Percent) Resulting in Two Positive Tests |
		Trial #1	Trial #2	Trial #3	Trial #4	Mean	
M	1	13	18	26	22	20	0
M,Th	2	34	32	47	48	41	0
M,W,F	3	61	53	67	66	63	0
M,W,Th,F	4	68	59	75	74	69	13.8
M,T,W,Th,F	5	77	69	81	81	79	26.3
M,T,W,Th,F,Sa	6	93	89	95	95	93	33.0
M,T,W,Th,F,Sa,S	7	100	100	100	100	100	48.3

*Each trial consists of 100 randomly generated times during the week that a person might self-administer a single dose of cocaine. A detection time of 35.8 hours was used in the determination of %drug episodes detected.

(one use went undetected) were obtained between 43 to 60 percent of the time, and double positive results (both uses were detected) were obtained at a frequency of 27 to 45 percent. When the single and double positive results are combined, the efficiency of detection of cocaine use for the week averaged 87.3 percent across the different drug use patterns. There were a maximum of seven instances of drug use occurring in the same detection time window when the second drug use could occur within 6 hours of the first use. In these instances, two uses appeared as a single use from the testing result. As the drug use interval lengthened to 24 hours, this phenomenon disappeared and was no longer a problem.

The data shown in tables 1 and 2 were generated to challenge the earlier conclusion that a 3-days-per-week schedule was the best compromise between maximizing drug detection and minimizing carryover. A Monday, Wednesday, Friday testing schedule demonstrated a mean efficiency of 63 percent in detecting single incidents of cocaine use. The increase in efficiency by further testing was relatively minimal until the frequency was increased to 6 days or more per week. Carryover of drug use from one test to another was not a factor with the Monday, Wednesday, Friday testing schedule but did occur at higher frequency testing schedules. When multiple cocaine use was simulated, that is, 2-times-per-week separated by a minimum time interval, the 6-days-per-week testing schedule was substantially better than a 1-day-per-week schedule.

TABLE 2. *Effect of urinalysis testing schedules on detection of two cocaine uses separated by a minimum hourly interval between uses during a week of testing**

Urinalysis Testing Schedule	Minimum Hours Between Drug Use	Number of Positive Test Results			Number of Times That Two Occurred in the Same Detection Period and Were Counted as One Positive Test
		0	1	2	
M	6	64	36	—	3
	12	67	33	—	3
	18	59	41	—	1
	24	50	50	—	1
	36	51	49	—	0
	48	48	52	—	0
	72	47	53	—	0
	84	39	61	—	0
M,W,F	6	16	50	34	7
	12	13	60	27	5
	18	13	51	36	3
	24	6	49	45	0
	36	11	47	42	0
	48	12	43	45	0
	72	15	43	42	0
	84	16	43	41	0

*Each trial consisted of 100 randomly generated time pairs (separated by a minimum interval) during the week that a person might self-administer two single doses of cocaine. A detection time of 35.8 hours was used in the determination of number of positive results.

SUMMARY AND CONCLUSIONS

Urinalysis can be used as an objective criterion for monitoring the outcome of a treatment program or a clinical trial. Important factors to consider when implementing a drug testing program include standardization of assay technology and cutoffs between participating centers and selection of identical testing schedules. Also, it is vitally important to minimize the amount of safe time (time that drug use can go undetected) occurring in a testing schedule. The detection times for cocaine and heroin have been shown to vary with selection of cutoff and with the drug dose. Obviously, the selection of cutoffs is under program control, whereas the amount of illicit drug use is under subject control. Fortunately, changes in the illicit drug dose by the subject demonstrate a log-linear relationship to detection time. Hence, a higher drug dose by the

subject only extends the detection time slightly (and improves the probability of detection) without greatly increasing the risks of drug carryover from one urine test to another.

The most efficient testing schedule for judging the outcome of clinical trials for cocaine and heroin appears to be a 3-days-a-week schedule (Monday, Wednesday, Friday or Tuesday, Thursday, Saturday). When different schedules were challenged by simulating random times at which cocaine use might occur during the week, the 3-days-per-week schedule was the most efficient without the risk of carryover. The 3-days-per-week schedule also performed better than 1-day-per-week when multiple random drug use was simulated. Overall, the 3-days-per-week testing schedule with specified assay technology and cutoffs was the best compromise for maximizing detection of drug use, minimizing carryover, and providing a standardized methodology for outcome comparison between programs.

REFERENCES

Cone, E.J.; Dickerson, S.; Paul, B.D.; and Mitchell, J.M. Forensic drug testing for opiates: IV. Analytical sensitivity, specificity and accuracy of commercial urine opiate immunoassays. *J Anal Toxicol* 16:72-78, 1992.

Cone, E.J.; Menchen, S.L.; Paul, B.D.; Mell, L.D.; and Mitchell, J. Validity testing of commercial urine cocaine metabolite assays: I. Assay detection times, individual excretion patterns, and kinetics after cocaine administration to humans. *J Forensic Sci* 34:15-31, 1989.

Cone, E.J., and Mitchell, J. Validity testing of commercial urine cocaine metabolite assays: II. Sensitivity, specificity, accuracy and confirmation by gas chromatography/mass spectrometry. *J Forensic Sci* 34:32-45, 1989.

Goldstein, A., and Brown, B.W., Jr. Urine testing schedules in methadone maintenance treatment of heroin addiction. *JAMA* 214:314-315, 1970.

Gorodetzky, C.W. Detection of drugs of abuse in biological fluids. In: Born, G.V.R.; Eichler, O.; Farah, A.; Herken, H.; and Welch, A.D., eds. *Handbook of Experimental Pharmacology.* Vol. 45. Berlin: Springer-Verlag, 1977. pp. 319-409.

Harford, R.J., and Kleber, H.D. Comparative validity of random-interval and fixed-interval urinalysis schedules. *Arch Gen Psychiatry* 35:356-359, 1978.

Mandatory guidelines for Federal workplace drug testing programs; final guidelines; notice. *Federal Register* 53:11970-11989, Apr. 11, 1988.

ACKNOWLEDGMENT

AUTHORS

Edward J. Cone, Ph.D.
Chief
Laboratory of Chemistry and Drug Metabolism

Sandra L. Dickerson, B.S.
Medical Technologist

Addiction Research Center
National Institute on Drug Abuse
P.O. Box 5180
Baltimore, MD 21224

Comments on "Efficacy of Urinalysis in Monitoring Heroin and Cocaine Abuse Patterns: Implications in Clinical Trials for Treatment of Drug Dependence" by Cone and Dickerson

Nancy L. Geller

Cone and Dickerson consider fixed-interval scheduling for drug use monitoring in trials for treatment of drug dependence. They conclude that changes in drug dose by the user alter detectability of drugs by urinalysis only slightly and that the Monday, Wednesday, Friday monitoring schedule is optimal because it maximizes the chance of detection of an episode of drug use and minimizes the chance of having two detections of the same episode.

The conclusion that dose alters detectability only slightly assumes a log-linear relationship between drug dose and detection times. This is equivalent to a one-compartment pharmacokinetic model. The data for morphine and heroin in Cone and Dickerson's figure 3 (this volume) suggest that a higher order compartmental model might be more appropriate. Such a possibility should be investigated.

The authors' conclusion that the Monday, Wednesday, Friday test schedule is optimal rests on certain assumptions:

1. If there is any episode of drug use, the test schedule should be able to detect it most of the time.

2. Detection of drug use within approximately 36 hours of that use is certain; that is, there are no false negatives.

3. Having two tests detect one episode of drug use should be avoided if possible.

TABLE 1. *Effect of urinalysis testing schedules on detection of a single random episode of cocaine use during a week of testing**

Urinalysis Testing Schedule	Simulated Probability of Detection of Drug Episode (n=400)	Actual Probability of Detection of Drug Episode	Simulated Probability of Drug Episode Resulting in Two Positive Tests (n=400)	Actual Probability of Drug Episode Resulting in Two Positive Tests
None	0	0	0	0
M	.20	.213	0	0
M,Th	.41	.426	0	0
M,W,F	.63	.639	0	0
M,W,TH,F	.69	.712	.138	.140
M,T,W,Th,F	.79	.785	.263	.281
M,T,W,Th,F,Sa	.93	.927	.330	.351
Every day	1.00	1.00	.483	.492

*A detection time of 35.8 hours and a zero probability of false negative tests were assumed in both the simulations and calculations.

4. If drug use is detected, there has indeed been drug use; that is, there are no false positives.

5. Drug detection will be done in multiples of 24 hours,

The probabilities that are simulated, according to the assumptions above, can be calculated exactly and are shown in table 1. As in the simulations, the exact calculations assumed that an episode of drug use is equally likely to occur at any time during the week (i.e., uniformly distributed). However, a trial participant who is going to take the drug may recognize that he or she is less likely to test positive next time if the drug is used soon after a urine test. Similarly, the probabilities of the model assumed for Cone and Dickerson's table 2 (this volume) can be explicitly calculated, but again, the times of an episode of drug taking may not be uniformly distributed.

Simulation is a rich tool and could allow more complicated scenarios to be evaluated, including nonuniform times of drug use. The possibility of false positives and false negatives could be built into a simulation model, which is equivalent to varying the cutoff for detection from 300 ng/mL. Testing at more than one time of day, such as mornings or afternoons, could also be evaluated. Software for simulating stochastic processes, such as the General Purpose Simulation System, might be used so that, in addition, random test times could be assessed.

The conclusions in Cone and Dickerson's chapter follow logically from their assumptions. However, more complex assumptions might be more realistic and could be considered in further work.

ACKNOWLEDGMENT

AUTHOR

Nancy L. Geller, Ph.D.
Chief
Biostatistics Research Branch
National Heart, Lung, and Blood Institute
National Institutes of Health
Federal Building, Room 2A-11
7550 Wisconsin Avenue
Bethesda, MD 20892

Summary of Discussion: "Efficacy of Urinalysis in Monitoring Heroin and Cocaine Abuse Patterns: Implications in Clinical Trials for Treatment of Drug Dependence" by Cone and Dickerson

Ram B. Jain

Dr. Weng suggested that blood samples from each subject be obtained prior to entry into the trial so that their individual pharmacokinetic profiles could be studied and their metabolic rates evaluated. The differences in metabolic rates will have a bearing on the detectability of drugs in urine. Individual pharmacokinetic profiles could also be used to appropriately schedule collection of urine samples. This suggestion was appreciated; however, as pointed out by Dr. Johnson, it is not practical to obtain blood samples from every subject, since some have poor venous access due to abuse of their veins from frequent injections. In addition, the process of randomization should equally distribute fast and slow metabolizers across different treatment groups,

Dr. Wright inquired about the cross-reactivity between opiates of abuse and replacement (treatment) opiates (and over-the-counter drugs) in immunoassays and about the need for confirmatory testing. According to Dr. Cone, the probability of false positives in immunoassays to detect opiates is very small unless a subject is using codeine. Use of a confirmatory assay such as gas chromatography/mass spectrometry would add little unless there was a need for quantitative data.

Dr. Gorodetzky asked if, in testing an individual by immunoassay following drug usage, negative results could be followed by positive results. It was acknowledged that this does not happen very often except with marijuana.

Dr. Fisher suggested that urine specimens be collected every day to collect the maximum amount of information. He suggested that this information could then be used to more appropriately interpret and/or modify information obtained from

Monday, Wednesday, and Friday specimens. He also suggested the need for estimating the amount of opiates used by using a method such as area under the curve. This method is probably impractical since many urine specimens would have to be collected over time, or timed plasma specimens with knowledge of duration since injection and amount of drug injected would be required. Dr. Johnson said different subjects may need different amounts of opiates to have the same effect, and because of risk of human immunodeficiency virus infection from intravenous injection of drugs using shared needles, it is important to know the exposure frequency and what the treatment drug can do to reduce this frequency.

Dr. Gordon also proposed to collect urine specimens more often than three times a week and, based *on* the results of a certain number of successive specimens (e.g., positive, negative, positive, positive), develop an algorithm to decide whether two or more consecutive positive specimens represent independent episodes of drug abuse or carryover. The proposal was well taken, but the same algorithm cannot be applied to all subjects since the probability of carryover varies from subject to subject. Such an algorithm has the potential to underestimate the probability of drug abuse. However, such an algorithm used in conjunction with self-reported drug use might be a possibility.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville. MD 20857

Open/Panel Discussion: Design Issues

Ram B. Jain

Panel Members: A.S. Hedayat (Chair), Albert J. Getson, Alan J. Gross, Sudhlr Gupta, Don Jasinski, Mel-Ling Ting Lee, Carol K. Redmond, and Margaret Wu

The three primary issues discussed were:

- Fixed time vs. random time sampling, including sampling frequency
- Estimation of carryover
- Estimation of the amount of drug abuse

FIXED TIME VS. RANDOM TIME SAMPLING

It was opined that the objectives of the clinical trials would determine the adequacy of fixed or random time sampling. If the objective was merely to evaluate the efficacy of a treatment drug, fixed time sampling would probably be the sampling scheme of choice. If determination of the effectiveness of the treatment drug was the objective of the trial, then random time sampling would probably be the sampling scheme of choice.

It was pointed out that determination of pharmacological efficacy of a treatment drug was the primary objective of a clinical trial such as the ARC 090 trial completed at the Addiction Research Center. The pharmacological efficacy was primarily evaluated by posttreatment frequency of drug abuse. However, as pointed out by Dr. Johnson, variables such as retention rates, withdrawal symptoms and signs, and opiate- and cocaine-craving scores were also evaluated in the ARC 090 trial.

Since the frequency of drug abuse is not directly measurable, frequency of detected drug (ab)use from urine samples on a per-sample or per-week basis is the surrogate measure used to represent posttreatment frequency of drug abuse. Using this surrogate measure, it is possible that multiple episodes of drug abuse are counted as one, but this is the limitation of the sampling

techniques currently used. In addition to reduction in the frequency of drug abuse, as Dr. Vocci mentioned, there is also interest in knowing when the treatment drug starts working. Some individuals in the buprenorphine and methadone 60 mg arms of the ARC 090 study stopped abusing drugs almost immediately and remained drug-free throughout the trial. Some individuals need to build a reservoir of treatment drug in the body before the drug shows its effect; it takes these individuals some time (4 to 6 weeks) before they stop abusing drugs. Eventually, a receptor occupancy *may* be reached for all individuals that may be consistent with no more drug (ab)use.

Since agonists, partial agonists, and antagonists act differently, there is also interest in evaluating the pattern of cessation of drug abuse, that is, the pattern of positive and negative urine samples. There is interest in being able to know if some daily users are being converted to weekend users only, for example, the maximum duration for which they can remain drug-free. For example, one of the outcome variables analyzed for the ARC 090 trial was the time to the (first) drug-free period of 28 days or more as determined by negative urines.

However, as pointed out by Drs. Wright and Getson, it is critical to remember that reduction in frequency and/or amount of drug abuse is only a small part of the claims that can be made for a treatment drug. These medications may also alter symptoms of drug (ab)use; suffering from drug (ab)use; social functioning (behavior) such as employment stability, family life, and crime-related activities; or target behaviors such as needle sharing and injection of illicit drugs. Evaluation of efficacy should be married to the development of the treatment compound as a whole. Efficacy trials should be followed by *effectiveness trials,* which may focus more on the sociological behaviors, as mentioned earlier. For these *effectiveness trials,* random time sampling may be the sampling scheme of choice. These *effectiveness trials* should lead to a broader understanding of the compound as a whole. An *efficacy trial* determines whether or not the drug works; an *effectiveness trial* generates additional information helpful in writing a good label (package insert) for the treatment drug.

An efficacious drug in the hands of a good clinician would work more effectively since these clinicians are likely to supplement treatment drugs with services such as family and/or employment counseling. However, if these *effectiveness variables* are allowed to interact with *efficacy variables* in the *efficacy trials,* the sample size requirements would become prohibitive and it may not be possible to show the pharmacological *efficacy* of the treatment drugs. It was suggested that *efficacy trials* may include an additional treatment arm in which subjects get other services, such as counseling, after only 2, 4, or 8 weeks of drug therapy.

If frequency of detectable drug abuse is the primary outcome variable in an efficacy trial, the drug abuse phenomenon should be viewed on a continuum, and as such, fixed time sampling should be appropriate. In fact, Dr. Fisher strongly favored collection of urine samples more often than three times a week, probably every day, since more information is generally better. However, collecting urine samples too often may have a negative effect on dropout rates and will shape the patient population remaining in the trial in such a way that generalizations to the addict population-at-large may be difficult. In fact, dropout rates are substantial (as much as 60 to 80 percent) in these trials. Cost may be another factor that should be considered. A compromise may be to do less frequent sampling in those who remain in the trial for a certain period and get as much information as possible on those who dropped out by sending out nurses, social workers, etc. There was a strong feeling that additional information should be obtained on those who drop out of the trial because, with dropout rates as high as they are in these trials, there certainly is a serious problem in making inferences for the total addict population. Also, such high dropout rates make it difficult to do an intent-to-treat analysis.

Dr. Jasinski pointed out that clinical trials are unique experiments as opposed to drug treatment programs. The lack of resources (financial and others) and practical considerations such as frequent collection of urine samples, if desired, should not stand in the way of doing these experiments. Resources should be obtained and study centers identified where these experiments can be successfully   conducted.

There were other arguments in favor of and against both fixed and random time sampling. It was suggested that fixed time sampling results in nonrandom missed observations. Since missing at random may be an assumption required to do some analyses, this may create a potential bias in these analyses. However, even in random time sampling, addicts are able to determine how often they will be tested and when, and thus, even random time sampling cannot ensure random missed observations. The data do not exist to show which type of sampling leads to higher noncompliance, including dropout rates. It may be that it is just the frequency (e.g., once a week vs. five times a week) of urine collection, irrespective of the type of sampling, that has a bearing on the noncompliance problem. In fixed time sampling, staffing requirements (to collect urine samples) are known in advance, which helps in planning for resources. A Food and Drug Administration audit of a trial done using fixed time sampling is relatively easier to conduct. Also, the choice between fixed time and random time sampling may be a choice between dealing with a possible treatment by day interaction and a relatively large error term (noise). Fixed time sampling may be used to collect data from some experimental units, and random sampling may be used to collect data from other experimental

units. But analysis of these data may present unknown challenges and possible interpretation problems. Alternatively, the data may be collected frequently using fixed time sampling, and only randomly selected data points may be used for analyses, In addition, irrespective of the type of sampling used, a clinical trial that has the ability to test (internally validate) some of the assumptions used in the analyses is preferred over one that does not have such an ability.

It was brought to attention that, since *efficacy trials* do not have any negative contingencies associated with results of urine samples, the question of whether data are missing at random may be a nonquestion since addicts may not have a reason to miss clinic visits. Dr. Blaine agreed. He explained that in their gepirone study, which did not have any negative contingencies associated with urine results, patients' admission of drug (ab)use matched urine test results most of the time. However, he added that absence of negative contingencies amounts to permission for drug abuse, as can be seen from the substantially higher percentage of positive urines (60 to 70 percent) from one of their buprenorphine trials, which did not have negative contingencies, compared with some treatment clinics (10 to 12 percent positive urines), which do have negative contingencies. Dr. Vocci emphasized that "if you are looking for efficacy in a clinical trial, you are better off . . . allowing individuals to use [drugs] in a manner that is not proscribed by the policies of the clinic." Artificially controlling drug abuse may result in prohibitive sample size requirements if a pharmacological effect is to be shown.

It was also suggested that the question of fixed vs. random time sampling should be decided by simulation methods using known pharmacokinetic profiles of the drugs that the urine samples are supposed to detect. These simulation methods may allow for a permissible degree of carryover and an inability to detect episodes of drug abuse. However, since pharmacokinetic profiles of the drugs of abuse are dose dependent and the dose and timing of drugs consumed by an addict are not known, such an exercise may be very difficult.

ESTIMATION OF CARRYOVER

It was mentioned that, in addition to a parallel design, a crossover design should be considered. A crossover design may be able to better handle the problem of carryover. However, as Dr. Hedayat pointed out, crossover designs have their own problems in interpretation of results. Parallel designs may be used to answer certain questions, whereas crossover trials may be designed to answer other questions.

Dr. Mei-Ling Lee visualized the problem in a different way. She observed that in these trials researchers are working with a mixture of distributions, and data should be analyzed as a mixing distributions problem. However, to analyze these data as a mixture of distributions, an estimate of the amount of drug present in the urine or plasma samples will be needed. Dr. Follmann agreed. Binary data would not be sufficient. Dr. Collins commented that, unless the timing of the episodes of drug abuse is known, concentration profiles of drug in the urine or plasma samples may not be fully informative. And as such, information obtained from pharmacokinetic profiles may have to be supplemented with that obtained by asking the addicts about the timing of drug abuse, if any, each time he or she is asked to provide a urine or plasma sample.

If binary data must be used, information obtained from a self-reported measure of drug abuse (e.g., "Did you use the drug during the last 24 hours?") when combined with the urine results may be able to help decide if a carryover existed. Dr. Wright suggested we should be looking for other sources of information, such as the staff and clinicians present at the time of clinic visits, The clinic staff should be able to judge if the subject may or may not have abused the drugs since the last clinic visit or since the last time a urine specimen was provided. Various pieces of information from various sources, including urine test results, can be put together according to a certain predefined set of rules and converted into some sort of scores that may be interpreted as a new episode of drug abuse or a carryover from the previous episode.

ESTIMATION OF THE AMOUNT OF DRUG ABUSE

A reduction in the frequency of drug abuse does not guarantee reduction in the amount of drug abuse. A daily user may be converted to an occasional user, but he or she may still be using the same amount of drug. Instead of using the same amount in, for example, 10 episodes, he or she might be using the same amount in 5 episodes. However, the binary data currently obtainable from urine assays cannot provide estimates of the amount of drug abuse. Hence, it was of interest to discuss this issue of being able to estimate the amount of drug abuse.

There was a strong sentiment at the meeting against attempts to estimate drug consumed from the drug present in the urine samples. As Dr. Gorodetzky put it, "Do not ask too much from a qualitative urine. You can be as precise as you want in terms of quantity of morphine in a . . . urine sample. It is not going to tell you a. thing about how much drug was taken, when and how many times. You cannot do it . . . There are some theoretical models you can build,

if you knew what time, when the time of drug administration was, if you knew when patients urinated, what the time was since last urination, what the volume was in this timed collection. Then maybe, if you had good enough data on which to base it, you could make some inferences. Right now, it cannot be done." Dr. Jasinski expressed similar sentiments.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical   Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

A Bayesian Nonparametric Approach to Analysis of Treatment for Drug-Dependence Data

Ram C. Tiwari

INTRODUCTION

In the National Institute on Drug Abuse ARC 090 trial to evaluate the efficacy of buprenorphine as compared with methadone 20 mg, and methadone 60 mg for treatment of opiate dependence, urine samples were obtained from patients three times (Monday, Wednesday, and Friday) every week for a period of 25 weeks. During the first 17 weeks of the study, the patients were maintained on the treatment drug; during the rest of the study, they were *detoxified* from the treatment drug. The urine samples were assayed for the presence of opiates. This chapter analyzes the data set collected from the first 17 weeks. As each dose of a treatment drug provides one observation, a maximum of 51 urine samples were obtained from each patient. The data also contain some missing observations due to no-shows during the course of study or due to withdrawals from the study. The accommodation of missing observations is an important issue and so is the use of information from the withdrawals from the study. To accommodate some missing observations, we have reduced 5 1 -dimensional data to 17-dimensional data by developing a weekly index of urine samples being positive or negative (see, also, Jain, this volume). A week is considered to be negative for opiates if at least two observations in this week are negative. Otherwise, the week is considered to be positive for opiates. Thus, the weeks with censored observations and two or more missing observations are automatically considered to be positive. This assumption does result in some loss of information, e.g., one who has three negative urines during a week is treated the same way as one who has only two negative urines during the week.

The next section presents a Bayesian approach to analysis of the binary response data. The analysis of ARC 090 trial data is presented in the last section.

BAYES ESTIMATION

Here, we consider a Bayesian nonparametric approach to the estimation of the conditional probabilities of the binary responses. To simplify the notations, denote a typical point $(t_1, \ldots, t_r)$ of the product space $T_r = \{0,1\}^r$ by $\underline{t}_r$ ($r = 1, 2, \ldots$). By $\underline{t}_r 0$ denote the point in T_{r+1} that is obtained by augmenting $\underline{t}_r$ by 0, that is, $\underline{t}_r 0 = (t_1, \ldots, t_r, 0)$, and similarly for $\underline{t}_r 1$. Finally, denote by $[\underline{t}_r]$ the cylinder set of all points in $T = \{0,1\}^\infty$ whose first r coordinates form the vector $\underline{t}_r$, that is, $[\underline{t}_r] = \{\underline{s} = (t_1, t_2, \ldots) \in T : \underline{s}_r = \underline{t}_r\}$.

Let $\mathcal{F}_r$ be the collection of the empty set and the finite disjoint unions of cylinder sets $[\underline{t}_r]$, $\underline{t}_r \in T_r$ $r = 1, \ldots$. The $\mathcal{F}_r$'s form an increasing sequence of σ-fields, and $\mathcal{F}^* = \bigcup_{r=1}^\infty \mathcal{F}_r$ is a field. The σ-field $\mathcal{F}$ in T is the smallest σ-field containing $\mathcal{F}^*$.

Consider a sequence of blocks ω of numbers in the closed unit interval $[0,1]$:

$$\omega = \{\pi, (\pi_1, \pi_0), (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00}), \ldots\} \tag{1}$$

For $t \in \{0,1\}$, let $\pi_{t_-}^t = \pi_{t_-}$ if $t = 1$, and $\pi_{t_-}^t = 1 - \pi_{t_-}$ if $t = 0$. Define a probability measure on P_r on $(T_r, \mathcal{F}_r)$ such that

$$
\begin{aligned}
P_r([\underline{t}_r]) &= \left(\prod_{j=1:t_j=1}^{r} \pi_{t_1,\ldots,t_{j-1}} \right) \left(\prod_{j=1:t_j=0}^{r} (1 - \pi_{t_1,\ldots,t_{j-1}}) \right), \\
&= \pi^{t_1} \pi_{\underline{t}_1}^{t_2} \pi_{\underline{t}_2}^{t_3} \ldots \pi_{\underline{t}_{r-1}}^{t_r}, \quad \underline{t}_r \in \mathcal{F}_r, \; r = 1, 2, \ldots. \tag{2}
\end{aligned}
$$

(Here and throughout t_0 is interpreted as the empty sequence, ϕ, and π_{t_0} is interpreted as π_-.) Then, it can be checked that

$$P_{r+1}([\underline{t}_r 1]) + P_{r+1}([\underline{t}_r 0]) = P_{r+1}([\underline{t}_r]) = P_r([\underline{t}_r]), \quad \underline{t}_r \in \mathcal{F}_r, \quad r = 1, \ldots.$$

Thus, the restriction of P_{r+1} to $\mathcal{F}_r$ is P_r, and P_r uniquely extends to a probability measure P on $(T, \mathcal{F})$.

Let $\Omega = \{\omega\}$ be the space of all blocks ω with its coordinates lying in $[0,1]$. If Ω is equipped with the product σ-field, $\sigma(\Omega)$, then the map $\omega \to P_\omega$ defines a transition function from $(\Omega, \sigma(\Omega))$ into $(T, \mathcal{F})$. Consider a probability measure Q on $(\Omega, \sigma(\Omega))$ such that, under Q, the coordinates of ω are mutually independent with

$$
\begin{aligned}
P_\omega([1]) &= \pi \sim \mathrm{Beta}(\alpha([1]), \alpha([0])), \text{ and} \\
P_\omega([\underline{t}_r 1]|[\underline{t}_r]) &= \pi_{\underline{t}_r} \sim \mathrm{Beta}(\alpha([\underline{t}_r 1]), \alpha([\underline{t}_r 0])), \; \underline{t}_r \in T_r, \; r = 1, 2, \ldots, \tag{3}
\end{aligned}
$$

where $\text{Beta}(a, b)$ denotes a Beta distribution with parameters a and b, then, under Q, the random probability measure P_ω is said to have a Polya tree process (Ferguson 1974). The posterior distribution of a Polya tree process, given an observation, is also a Polya tree process and is obtained by updating one of the Beta distributions at each level of the tree.

If in (3)α is a finite measure on $(\mathcal{T}, \mathcal{F})$, then, under Q, the joint distribution of $\{P_\omega([\underline{t}_r]); \, \underline{t}_r \in T_r\}$ is a Dirichlet distribution with parameter $\{\alpha([\underline{t}_r]); \, \underline{t}_r \in T_r\}$, $r = 1, 2, \ldots$(see Basu and Tiwari 1982). Furthermore, if $\{A_1, A_2, \ldots, A_k\} \subset \mathcal{F}^*$ is a partition of $\mathcal{T}$, then $(P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$ have a Dirichlet distribution with parameters $(A_1), \alpha(A_2), \ldots, \alpha(A_k))$, since $A_j \in \mathcal{F}^*(j = 1, 2, \ldots, k)$ implies that $A_j \in \mathcal{F}_r(j = 1, \ldots, k)$, for some r $(r = 1, \ldots, n)$. The following result is useful.

Theorem 1. (Blackwell 1973). The random probability measure P_ω is a Dirichlet process on $(\mathcal{T}, \mathcal{F})$ with parameter α.

To prove theorem 1 it suffices to show that, for arbitrary measurable partition $A_1, A_2, \ldots, A_k \subset \mathcal{F}$ of $\mathcal{T}$, the distribution of $(P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$ is a Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$ (cf. Ferguson 1973). This follows from the following lemmas.

Lemma 1. Let A be an arbitrary set in $\mathcal{F}$. Then, under the assumption (3), the random variable P_ω has a $\text{Beta}(\alpha(A), \alpha(A^c))$ distribution.

Proof. Let $\mathcal{C} = \{ A \in \mathcal{F} : P_\omega(A) \text{ has a } \text{Beta}(\alpha(A), \alpha(A^c)) \text{ distribution}\}$. Then, by definition $\mathcal{C} \subset \mathcal{F}$. It is easy see that $\mathcal{C}$ contains $\mathcal{F}^*$. Also, $\mathcal{C}$ is a monotone class. To see this, let $\{A_n\}_{n=1}^\infty$ be an increasing sequence of sets in $\mathcal{C}$, and let $A = \bigcup_{n=1}^\infty A_n$. Then, clearly $A \in \mathcal{F}$, and by continuity from below $P_\omega(A_n) \uparrow P_\omega(A)$, for each $\omega \in \Omega$, as $n \to \infty$, which is a stronger result than the convergence in distribution. Again, since all finite moments of the random variable $P_\omega(A_n)$ converge to the corresponding moments of a random variable X, say, having $\text{Beta}(\alpha(A), \alpha(A^c))$ distribution. Hence, $A \in \mathcal{C}$. For a decreasing sequence of sets in $\mathcal{C}$, we argue in a similar way. Thus, is a monotone class containing the field $\mathcal{F}^*$, and hence $\mathcal{C}$ contains $\mathcal{F}$, the smallest σ-field containing $\mathcal{F}^*$.

Lemma 2. For any arbitrary finite partition $(A_1, A_2, \ldots, A_k)$ for $\mathcal{T}$ in $\mathcal{F}$-measurable sets, the random variables $(P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$ have the Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$.

Proof. By an approximation theorem (see, Billingsley 1979, Theorem 11.4,

p. 140) there exists a sequence $\{(A_{1n}, A_{2n}, \ldots, A_{kn})\}_{n=1}^{\infty}$ of partitions of T into $\mathcal{F}^*$ measurable sets such that $\alpha(A_j \triangle A_{jn}) \to 0$ as $n \to \infty$, $j = 1, 2, \ldots, k$. Also, from Lemma 1, the random variable $P_\omega(A_j \triangle A_{jn})$ has a $\text{Beta}(\alpha(A_j \triangle A_{jn}), \alpha(A_j^c \triangle A_{jn}))$ distribution, $j = 1, 2, \ldots, k$. Therefore, for $j = 1, 2, \ldots, k$, we have

$$|P_\omega(A_j) - P_\omega(A_{jn})| \le P_\omega(A_j \triangle A_{jn}) \to 0 \quad \text{wp1}[Q] \text{ as } n \to \infty,$$

since $E_Q \, P_\omega(A_j \triangle A_{jn}) = \frac{\alpha(A_j \triangle A_{jn})}{\alpha(T)} \to 0$ as $n \to \infty$. Thus, the random variables $(P_\omega(A_{1n}), P_\omega(A_{2n}), \ldots, P_\omega(A_k))$ converge in probability and hence in distribution to the random variables $(P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$. Since, for each n, the random variables $(P_\omega(A_{1n}), \ldots, P_\omega(A_{kn}))$ have a Dirichlet distribution with parameters $(\alpha(A_{1n}), \alpha(A_{2n}), \ldots, \alpha(A_{kn}))$, and all finite moments of $(P_\omega(A_{1n}), P_\omega(A_{2n}), \ldots, P_\omega(A_{kn}))$ converge to the corresponding moments of the random variables $(X_1, X_2, \ldots, X_k)$, say, having a Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$, it follows that $(P_\omega(A_1), P_\omega(A_2), \ldots, P_\omega(A_k))$ have a Dirichlet distribution with parameters $(\alpha(A_1), \alpha(A_2), \ldots, \alpha(A_k))$.

For more on the Dirichlet process, see Ferguson (1973, 1974) or a recent survey article by Ferguson and colleagues (1992).

The map $\underline{t} = (t_1, t_2, \ldots) \to \sum_{r=1}^{\infty} \frac{t_r}{2^r}$ from T into $[0, 1]$ induces a random probability measure on $[0, 1]$. If P_ω is a Polya tree process on T, then the induced random probability measure on $[0, 1]$ is also a Polya tree process. Furthermore, if

$$\sum_{r=1}^{\infty} \sum_{\underline{t}_r \in T_r} \text{Var}(\pi_{\underline{t}_r}) < \infty,$$

(4)

then the induced random distribution function on $[0, 1]$ is absolutely continuous w.p. $1[Q]$ (cf. Kraft 1964 and Métiviar 1971). If P_ω is a Dirichlet process on $(T, \mathcal{F})$ with parameter α, then (4) simplifies to

$$\sum_{r=1}^{\infty} \sum_{\underline{t}_r \in T_r} \frac{\alpha([\underline{t}_r 1]) \alpha([\underline{t}_r 0])}{\alpha([\underline{t}_r])^2 (\alpha([\underline{t}_r]) + 1)} < \infty.$$

(5)

Suppose there are m patients involved in the study on a treatment. Corresponding to the ith patient, let $\underline{y}^i = (y_1^i, y_2^i, \ldots, y_n^i)$ denote the vector of observations on the response variable y taking on only two values: 1 = presence, and 0 = absence of opiate. Thus, $\underline{y}^i \in T_n$ Let $\underline{y}_j^i = (y_1^i, \ldots, y_j^i)$.

73

Given P_ω, let $\underline{t}^1, \underline{t}^2, \ldots, \underline{t}^m$ be i.i.d. observations from P_ω. Then, the likelihood function of ω is

$$L(\omega|\underline{t}_n^1 = \underline{y}^1, \ldots, \underline{t}_n^m = \underline{y}^m)$$

$$= \prod_{i=1}^m \left\{ \left(\prod_{j=1:y_j^i=1}^n \pi_{y_1^i \cdots y_{j-1}^i} \right) \left(\prod_{j=1:y_j^i=1}^n \left(1 - \pi_{y_1^i \cdots y_{j-1}^i} \right) \right) \right\}$$

$$= \prod_{i=1}^m \pi^{y_1^i} \pi_{\underline{y}_1^i}^{y_2^i} \pi_{\underline{y}_2^i}^{y_3^i} \cdots \pi_{\underline{y}_{n-1}^i}^{y_n^i}$$

$$= \pi^{m([1])} (1 - \pi)^{m([0])} \pi_1^{m([1])} (1 - \pi_1)^{m([10])} \cdots$$

$$= \prod_{r=0}^{n-1} \prod_{\underline{t}_r \in \mathcal{T}_r} \pi_{\underline{t}_r}^{m([\underline{t}_r 1])} (1 - \pi_{\underline{t}_r})^{m([\underline{t}_r 0])}, \tag{6}$$

where $m([\underline{t}_r]) = \sum_{i=1}^m \delta_{y^i}([\underline{t}_r])$, and $m([\underline{t}_r 0]) = m - m([\underline{t}_r 1])$, $\underline{t}_r \in \mathcal{T}_r$, and δ_a is the degenerate measure at a.

Clearly, under $\mathcal{Q}$,

$$P_\omega([\underline{t}_r]) \sim \text{Beta}(\alpha([\underline{t}_r]), \alpha(\mathcal{T}) - \alpha([\underline{t}_r])), \quad \underline{t}_r \in \mathcal{T}_r, \ r = 1, 2, \ldots.$$

Furthermore, from (3) and (6) it can be easily checked that the coordinates of w are mutually independent a posteriori, and

$$\pi|_{\underline{t}_n^1 = \underline{y}^1, \ldots, \underline{t}_n^m = \underline{y}^m} \sim \text{Beta}(\alpha([1]) + m([1]), \alpha([0]) + m([0]))$$

and

$$\pi_{\underline{t}_r}|_{\underline{t}_n^1 = \underline{y}^1, \ldots, \underline{t}_n^m = \underline{y}^m} \sim \text{Beta}(\alpha([\underline{t}_r 1]) + m([\underline{t}_r 1]), \alpha([\underline{y}_r 0]) + m([\underline{y}_r 0])),$$
$$\underline{t}_r \in \mathcal{T}_r, \ r = 1, 2 \ldots$$

From (2), (3) and (7) it follows that if P_ω is a Polya tree process, then P_ω, given the data, is again a Polya tree process. In particular, if P_ω is a Dirichlet process with parameter α, then P_ω, given the data, is also a Dirichlet process on $(\mathcal{T}, \mathcal{F})$ with updated parameter $\alpha(\cdot) + m(\cdot)$. Also, from (7) under squared error loss, the Bayes estimators of the conditional probabilities are given by

$$\hat{\pi} = \frac{\alpha([1]) + m([1])}{\alpha(\mathcal{T}) + m},$$

and

$$\hat{\pi}_{t_r} = \frac{\alpha([\underline{t}_r 1]) + m([\underline{t}_r 1])}{\alpha([\underline{t}_r]) + m([\underline{t}_r])}$$

$$= \frac{\alpha([\underline{t}_r 1])}{\alpha([\underline{t}_r]) + m([\underline{t}_r])} \left(\frac{\alpha([\underline{t}_r 1])}{\alpha([\underline{t}_r])} \right) + \frac{m([\underline{t}_r])}{\alpha([\underline{t}_r]) + m([\underline{t}_r])} \left(\frac{m([\underline{t}_r 1])}{m([\underline{t}_r])} \right) ,(8)$$

where $\underline{t}_r \in \mathcal{T}_r$.

The posterior variance of the conditional probabilities are given by

$$\mathrm{Var}_Q(\pi_{\underline{t}_r} | \underline{t}_n^1 = \underline{y}^1, \dots, \underline{t}_n^m = \underline{y}^m) = \frac{(\alpha([\underline{t}_r 1]) + m([\underline{t}_r 1]))(\alpha([\underline{t}_r 0]) + m([\underline{t}_r 0]))}{(\alpha([\underline{t}_r]) + m([\underline{t}_r]))^2 (\alpha([\underline{t}_r]) + m([\underline{t}_r]) + 1)}.$$

Also, the Bayes estimator of the unconditional probabilities $P_\omega([\underline{t}_r]), \underline{t}_r \in \mathcal{T}_r$ and their posterior variances are given by

$$\hat{P}([\underline{t}_r]) = E_Q(P_\omega([\underline{t}_r]) | \underline{t}_n^1 = \underline{y}^1, \dots, \underline{t}_n^m = \underline{y}^m) = \frac{\alpha([\underline{t}_r]) + m([\underline{t}_r])}{\alpha(T) + m}$$

$$= \frac{\alpha(T)}{\alpha(T) + m} \left(\frac{\alpha([\underline{t}_r])}{\alpha(T)} \right) + \frac{m}{\alpha(T) + m} \left(\frac{m([\underline{t}_r])}{m} \right), \qquad (9)$$

and

$$\mathrm{Var}_Q(P_\omega([\underline{t}_r]) | \underline{t}_n^1 = \underline{y}^1, \dots, \underline{t}_n^m = \underline{y}^m) = \frac{(\alpha([\underline{t}_r]) + m([\underline{t}_r]))(\alpha(T) + m - m([\underline{t}_r]))}{(\alpha(T) + m)^2 (\alpha(T) + m + 1)},$$

respectively.

ANALYSIS OF ARC 090 TRIAL DATA

As mentioned earlier, we have reduced 51-dimensional data to 17-dimensional data by developing a weekly index of urine samples being positive or negative. A week is considered to be negative for opiates if at least two observations in this week are negative. Otherwise, the week is considered to be positive for opiates. Clearly, this approach takes into account the censored observations, We have denoted the positive weeks by 1's and the negative weeks by 0's. For simplicity, we assume that the parameter α of the Dirichlet process is given by $\alpha([\underline{t}_r]) = \frac{1}{2^r}, \underline{t}_r \in \mathcal{F}_r.$ This corresponds to the Lebesgue measure on [0, 1]. Thus, for no sample case, the prior guess of the unconditional probability $P([\underline{t}_r])$ is $\frac{1}{2^r}$ and that of the conditional probability $\pi_{\underline{t}_r} = P([\underline{t}_r 1] | [\underline{t}_r])$ is $\frac{1}{2}$. The corresponding Bayes estimates $\hat{P}([\underline{t}_r])$ and $\hat{\pi}_{\underline{t}_r}$, for some selected sequences

75

for the three treatments buprenorphine, methadone 20 mg, and methadone 60 mg are given by the columns 2 and 3 in tables 1, 2, and 3. For example, if $r = 5$ and $[t_r] = [11111]$, then from table 1 we observe that $\hat{P}([t_r]) = 0.37094907407$ and $\hat{\pi}_{t_r} = \hat{P}([111111] \mid [11111]) = 0.9921996880$. Graphs of unconditional probabilities for some sequences (up to length five) for the three treatments are given in figures 1, 2, and 3.

TABLE 1. *Conditional and unconditional probabilities of some selected sequences for buprenorphine treatment*

t_r	$\hat{P}([t_r])$	$\hat{P}([t_r 1]\mid[t_r])$
1	7.1296296296E-01	7.0779220779E-01
11	5.0462962963E-01	8.8532110092E-01
10	2.0833333333E-01	2.7777777778E-01
00	2.6388888889E-01	1.4912280702E-01
111	4.4675925926E-01	8.3160621762E-01
100	1.5046296296E-01	3.7692307692E-01
1111	3.7152777778E-01	9.9844236760E-01
1110	7.5231481481E-02	9.9230769231E-01
1001	5.6712962963E-02	6.6326530612E-01
11111	3.7094907407E-01	9.9921996880E-01
11101	7.4652777778E-02	7.4806201550E-01
10011	3.7615740741E-02	9.9230769231E-01
00000	2.0428240741E-01	2.7337110482E-01
111111	3.7065972222E-01	9.4964871194E-01
111011	5.5844907407E-02	6.6580310881E-01
100000	7.4363425926E-02	5.0000000000E-01
100111	3.7326388889E-02	9.9612403101E-01
1111111	3.5199652778E-01	9.9979449240E-01
1110111	3.7181712963E-02	9.9805447471E-01
1000001	3.7181712963E-02	9.9805447471E-01
1001111	3.7181712963E-02	9.9805447471E-01
11111111	3.5192418981E-01	9.4727646454E-01
11101111	3.7109375000E-02	9.9902534113E-01
10011111	3.7109375000E-02	9.9902534113E-01
111111111	3.3336950231E-01	9.4439622437E-01
111011111	3.7073206018E-02	9.9951219512E-01
100111111	3.7073206018E-02	9.9951219512E-01
1111111111	3.1483289931E-01	9.4115112873E-01
1110111111	3.7055121528E-02	9.9975597853E-01
1001111111	3.7055121528E-02	9.9975597853E-01
11111111111	2.9630533854E-01	9.3748664897E-01
11101111111	3.7046079282E-02	9.9987795948E-01
10011111111	3.7046079282E-02	9.9987795948E-01
111111111111	2.7778229890E-01	9.9999186211E-01
111011111111	3.7041558160E-02	9.9993897229E-01
100111111111	3.7041558160E-02	9.9993897229E-01
1111111111111	2.7778003834E-01	9.9999593102E-01
1110111111111	3.7039297598E-02	9.9996948428E-01
1001111111111	3.7039297598E-02	9.9996948428E-01
11111111111111	2.7777890806E-01	9.9999796550E-01
11101111111111	3.7038167318E-02	9.9998474168E-01
10011111111111	3.7038167318E-02	9.9998474168E-01
111111111111111	2.7777834292E-01	9.9999898275E-01
111011111111111	3.7037602177E-02	9.9999237072E-01
100111111111111	3.7037602177E-02	9.9999237072E-01

The tables and figures show that for all the three treatments the probabilities for consecutive positive weeks of a fixed length are larger than the probabilities of any other sequences of the same length. The probabilities of consecutive positive weeks for buprenorphine and methadone 60 mg are smaller than the corresponding probabilities for methadone 20 mg.

TABLE 2.   *Conditional and unconditional probabilities of some selected sequences for methadone 20-mg treatment*

$\underline{t}_r$	$\hat{P}([\underline{t}_r])$	$\hat{P}([\underline{t}_r,1]|[\underline{t}_r])$
1	7.9464285714E-01	8.3707865168E-01
00	1.8303571429E-01	4.0243902439E-01
11	6.6517857143E-01	9.6979865772E-01
10	1.2946428571E-01	5.6896551724E-01
001	7.3660714286E-02	7.4242424242E-01
111	6.4508928571E-01	8.8754325259E-01
100	5.5803571429E-02	6.6000000000E-01
0011	5.4687500000E-02	6.6326530612E-01
1111	5.7254464286E-01	9.6783625731E-01
1110	7.2544642857E-02	7.4615384615E-01
1001	3.6830357143E-02	9.8484848485E-01
00111	3.6272321429E-02	9.9230769231E-01
00000	7.1986607143E-02	5.0000000000E-01
11111	5.5412946428E-01	9.9949647533E-01
11101	5.4129464286E-02	9.9484536082E-01
10100	5.4129464286E-02	9.9484536082E-01
10011	3.6272321429E-02	9.9230769231E-01
001111	3.5993303571E-02	9.9612403101E-01
000001	3.5993303571E-02	9.9612403101E-01
111111	5.5385044643E-01	9.9974811083E-01
111011	5.3850446429E-02	9.9740932642E-01
100111	3.5993303571E-02	9.9612403101E-01
0011111	3.5853794643E-02	9.9805447471E-01
1111111	5.5371093750E-01	9.9987402368E-01
1110111	5.3710937500E-02	6.6623376623E-01
00111111	3.5784040179E-02	9.9902534113E-01
11111111	5.5364118303E-01	9.9993700390E-01
11101111	3.5784040179E-02	9.9902534113E-01
001111111	3.5749162946E-02	9.9951219512E-01
111111111	5.5360630580E-01	9.9996849997E-01
111011111	3.5749162946E-02	9.9951219512E-01
0011111111	3.5731724330E-02	9.9975597853E-01
1111111111	5.5358886719E-01	9.6772720113E-01
1110111111	3.5731724330E-02	9.9975597853E-01
00111111111	3.5723005022E-02	9.9987795948E-01
11111111111	5.3572300502E-01	9.6665907130E-01
11101111111	3.5723005022E-02	9.9987795948E-01
001111111111	3.5718645368E-02	9.9993897229E-01
111111111111	5.1786150251E-01	9.9999579071E-01
111011111111	3.5718645368E-02	9.9993897229E-01
0011111111111	3.5716465541E-02	9.9996948428E-01
1111111111111	5.1785932268E-01	9.9999789535E-01
1110111111111	3.5716465541E-02	9.9996948428E-01
00111111111111	3.5715375628E-02	9.9998474168E-01
11111111111111	5.1785823277E-01	9.9999894767E-01
11101111111111	3.5715375628E-02	9.9998474168E-01
001111111111111	3.5714830671E-02	9.9999237072E-01
111111111111111	5.1785768781E-01	9.9999947383E-01
111011111111111	3.5714830671E-02	9.9999237072E-01

TABLE 3. *Conditional and unconditional probabilities of some selected sequences for methadone 60-mg treatment*

t_r	$\hat{P}([t_r])$	$\hat{P}([t_r 1]/[t_r])$
0	2.2727272727E-01	5.0000000000E-01
1	7.7272727273E-01	7.1176470588E-01
01	1.1363636364E-01	8.2000000000E-01
10	2.2272727273E-01	4.1836734694E-01
11	5.5000000000E-01	8.3057851240E-01
011	9.3181818182E-02	9.8780487805E-01
101	9.3181818182E-02	7.9268292683E-01
100	1.2954545455E-01	4.2982456140E-01
111	4.5681818182E-01	7.9850746269E-01
110	9.3181818182E-02	4.0243902439E-01
0111	9.2045454545E-02	7.9629629630E-01
1011	7.3863636364E-02	7.4615384615E-01
1001	5.5681818182E-02	6.6326530612E-01
1111	3.6477272727E-01	9.4859813084E-01
1110	9.2045454545E-02	4.0123456790E-01
1101	3.7500000000E-02	9.8484848485E-01
01111	7.3295454545E-02	9.6612403101E-01
10111	5.5113636364E-02	9.9484536082E-01
10011	3.6931818182E-02	9.9230769231E-01
11111	3.4602272727E-01	9.9917898194E-01
11100	5.5113636364E-02	6.6494845361E-01
11101	3.6931818182E-02	9.9230769231E-01
11011	3.6931818182E-02	9.9230769231E-01
011111	7.3011363636E-02	9.9805447471E-01
101111	5.4829545454E-02	9.9740932642E-01
100111	3.6647727273E-02	9.9612403101E-01
111111	3.4573863636E-01	9.4700082169E-01
0111111	7.2869318182E-02	9.9902534113E-01
1011111	5.4687500000E-02	9.9870129870E-01
1001111	3.6505681818E-02	9.9805447471E-01
1111111	3.2741477273E-01	9.4425162690E-01
01111111	7.2798295454E-02	7.4975609756E-01
10111111	5.4616477273E-02	9.9934980494E-01
10011111	3.6434659091E-02	9.9902534113E-01
11111111	3.0916193182E-01	9.9988513669E-01
011111111	5.4580965909E-02	9.9967469095E-01
101111111	5.4580965909E-02	9.9967469095E-01
100111111	3.6399147727E-02	9.9951219512E-01
111111111	3.0912642045E-01	9.9994256174E-01
0111111111	5.4563210227E-02	9.9983729255E-01
1011111111	5.4563210227E-02	9.9983729255E-01
1001111111	3.6381392045E-02	9.9975597853E-01
1111111111	3.0910866477E-01	9.9997127922E-01
01111111111	5.4554332386E-02	9.9991863303E-01
10111111111	5.4554332386E-02	9.9991863303E-01
10011111111	3.6372514204E-02	9.9987795948E-01
11111111111	3.0909978693E-01	9.9998563920E-01
011111111111	5.4549893466E-02	9.9995931321E-01
101111111111	5.4549893466E-02	9.9995931321E-01
100111111111	3.6368075284E-02	9.9993897229E-01
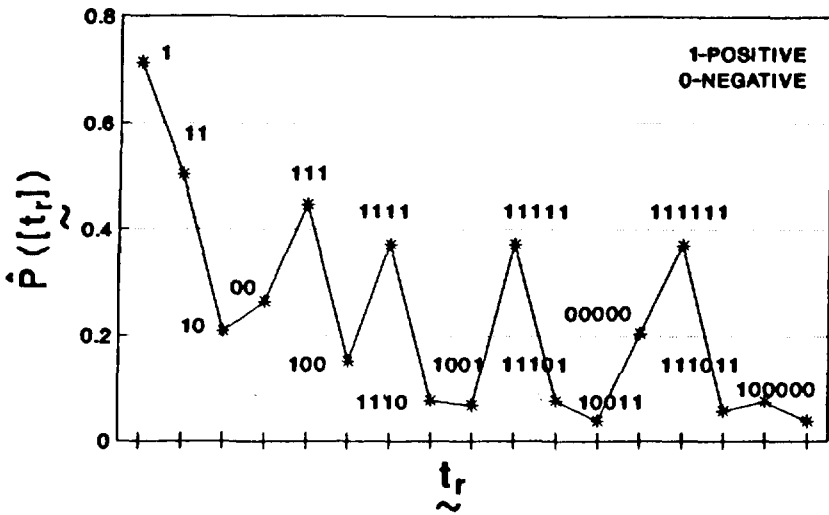
FIGURE 1. *Unconditional probabilities for buprenorphine treatment*
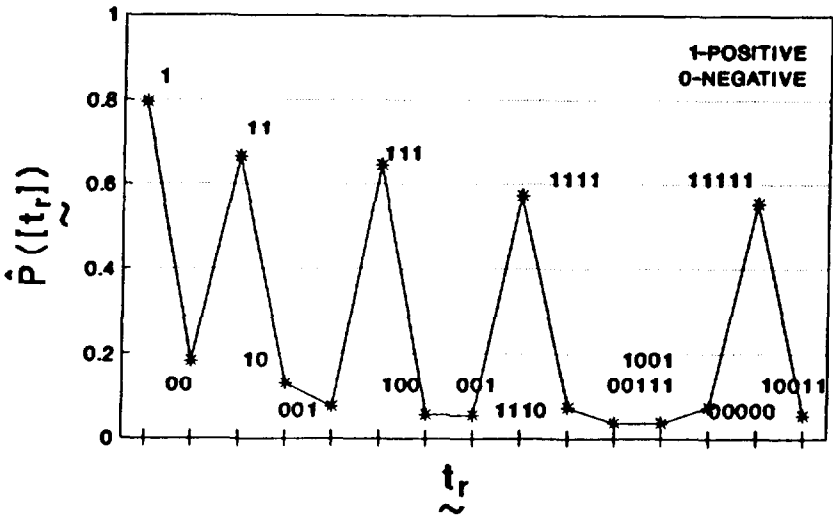
FIGURE 2. *Unconditional probabilities for methadone 20-mg treatment*
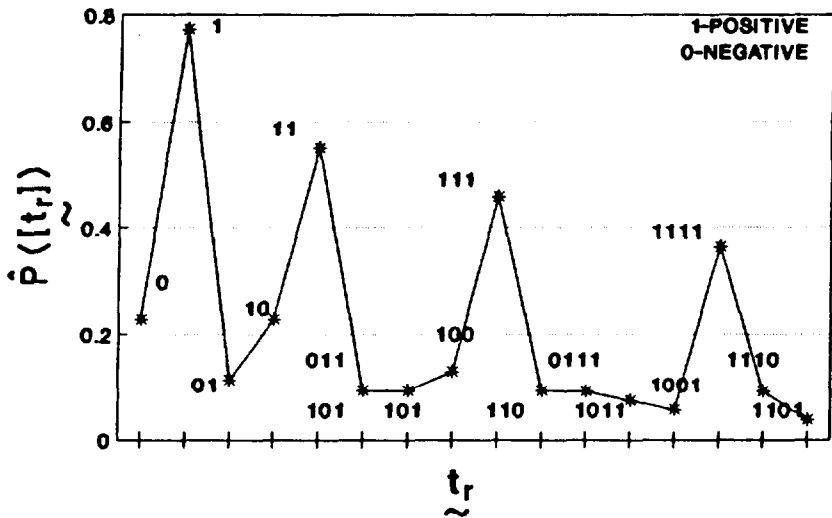
FIGURE 3. *Unconditional probabilities for methadone 60-mg treatment*

REFERENCES

Basu, D., and Tiwari, R.C. A note on the Dirichlet process. In: Kallianpur, G.; Krishnaiah, P.R.; Ghosh, J.K., eds. *Statistics and Probability: Essays in Honor of C.R. Rao.* New York: North-Holland, 1982. pp. 89-103.

Billingsley, P. *Probability and Measures.* New York: Wiley, 1979.

Blackwell, D. Discreteness of Ferguson selections. *Ann Stat* 1:356-358, 1973.

Ferguson, T.S. A Bayesian analysis of some nonparametric problems. *Ann Stat* 1:209-230, 1973.

Ferguson, T.S. Prior distributions on spaces of probability measures. *Ann Stat* 2:615-629, 1974.

Ferguson, T.S.; Phadia, E.G.; and Tiwari, R.C. Bayesian nonparametric inference. In: Ghosh, M., and Pathak, PK., eds. *Essays in Honor of D. Basu Vol. 17, IMS Lecture Notes - Monograph Series.* Hayward, CA: Institute of Mathematical Statistics, 1992.

Kraft, C.H. A class of distribution function processes which have derivatives. *J Appl Probability* 1:385-388, 1964.

Metivier, M. Sur la construction de mesures aléatoirs presque súrment absolument continues par rapport à une mesure donnée. *Z. Wahrscheinlichkeitstheorie view. Geb.* 20:332-344, 1971.

AUTHOR

Ram C. Tiwari, Ph.D.
Associate   Professor
Department of Mathematics
University of North Carolina at Charlotte
Charlotte, NC 28223

Three Estimators of the Probability of Opiate Use From Incomplete Data

Alan J. Gross

INTRODUCTION

Drug testing in biologic fluids, especially urine, has become the usual method by which addicts in a treatment program are evaluated to determine whether they are adhering to the treatment regime in which they have been placed. Unfortunately, issues such as sensitivity and specificity of the various tests have caused difficulties in the past. The Council on Science Affairs (1987) has dealt with these important issues.

Besides these issues of sensitivity and specificity, there are concerns about whether a random time or a *fixed* time sampling scheme should be used when collecting urine specimens from subjects who are involved in a clinical trial that is designed to test the safety and efficacy of a new pharmocotherapy for treatment of opiate dependence.

It is the purpose of this chapter to consider three estimators of the probability that an addict tests positive for a particular opiate and compare these estimates for the ARC 090 data that were generated by means of fixed time sampling. The properties of these estimators will also be investigated, and some preliminary results will be given. Although other important issues exist in this area of research, such as random time sampling schemes, they are not specifically addressed in this chapter.

ESTIMATORS OF THE PROBABILITY OF OPIATE USE

In an effort to estimate the probability of opiate use by an addict during the time period in which he or she is enrolled in a clinical trial designed to reduce drug dependence, the following assumptions and definitions are required.

Assume that an individual within a clinical trial is scheduled to present m times for testing of the presence of the opiate for which he or she is being treated. On

the ith visit, the random variables U_i and Δ_i are defined as

$$U_i = \begin{cases} 1 & \text{the individual tests positive for the opiate,} \\ 0 & \text{the individual tests negative} \end{cases}$$

and

the individual appears for the test and is still in the trial,
the individual does not appear

$i = 1, \ldots, m$. It is noted that m can and does change from subject to subject, and in the data set that is considered when treatment groups are compared, once a subject has been censored from the trial, that subject never returns for any future testing. It is further assumed that

(a) $\{U_i\}_{i=1}^m$ are Bernoulli random variables such that

(i) $\mathrm{pr}\{U_i = 1\} = p, 0 \le p \le 1$, and $\mathrm{corr}(U_i, U_j) = \rho_{ij}, 0 \le \rho_{ij} \le 1$, where in the first case

$$\rho_{ij} = \rho^{|i-j|}, \tag{1}$$

in the second case

(ii)
$$\rho_{ij} = \rho, \tag{2}$$

$i \ne j$, and in the third case, the correlation structure

(iii)
$$\rho_{ij} = \begin{cases} \rho & \text{if } j = i+1 \\ 0 & \text{if } j > i+1 \end{cases} \qquad \text{for } j > i, \quad i = 1, \ldots, m-1. \tag{3}$$

and

(b) $\{\Delta_i\}_{i=1}^m$ are iid Bernoulli random variables such that

$$\mathrm{pr}\{\Delta_i = 1\} = \pi, \qquad 0 \le \pi \le 1,$$

U_i and Δ_i, U_i and Δ_j, $i \ne j$ are assumed independent.

The three correlation structures considered here deal then, respectively, with the following three scenarios:

83

1. There is correlation between all pairs of visits for a given individual. In this case, the correlation between successive visits is assumed to be greater (in absolute value) than visits that are more distant. The structure is the same as in the simple autoregressive model.

2. The correlation in this case is assumed to stay constant between all pairs of visits, successive as well as more distant, within an individual. This assumption may tend to be somewhat conservative.

3. The correlation between successive visits within an individual is assumed and is constant within individuals. It is also assumed constant from individual to individual. However, more distant visits are assumed to be uncorrelated.

The dependence or correlation structure presented in this chapter differs, to some extent, from correlated binomial random variables that were considered in other applications. These earlier applications include correlated binomial models to predict the probability of rainfall on a given day realizing that the occurrence or nonoccurrence of rain on a given day depends on the occurrence or nonoccurrence on the previous day. Such models were developed by Gabriel (1959), Gabriel and Neuman (1962), and Klotz (1973). In the model considered by Klotz (1973), it can be shown that

$$\rho^{|i-j|} = [(p^* - p)(1 - p)]^{|i-j|} \quad \text{where } p* = \text{pr}\{X_k = 1 \mid X_{k-1} = 1\}.$$

As a second example in ophthalmology studies, a particular disease may be present in one eye, both eyes, or neither eye in a patient. Rosner (1984) considers a correlated binomial model in this situation because, clearly, absence or presence of disease in the two eyes of an individual is not independent from eye to eye. Finally, in this vein, Kupper and Haseman (1978) and Haseman and Kupper (1979) apply correlated binomial models to analyzing data within and among animal litters for which the responses are dichotomous, e.g., occurrence or nonoccurrence of a malformation.

Consider now, the estimator for the probability of opiate use. Define $\hat{P}$ as

$$\hat{P} = \begin{cases} \sum_{i=1}^{m} U_i \Delta_i / \sum_{i=1}^{m} \Delta_i, & \text{if } \sum_{i=1}^{m} \Delta_i \geq 1, \\ 1, & \text{if } \Delta_1 = \Delta_2 = \cdots = \Delta_m = 0. \end{cases} \tag{4}$$

This definition indicates that an individual who is never present for a test to determine his or her drug abuse status is very likely, if not certain, to be still

abusing the drug. It should be noted that (1) this definition does not distinguish all sequences, for example, it does not distinguish 000111 and 111000; and (2) $\hat{P}$ represents an average across visits for each individual. Although, these are somewhat limiting, it is an initial attempt to deal with such binomial data in a relatively simple manner.

The first principal goal of this chapter is to obtain $(\hat{P})$ and $\text{Var}(\hat{P})$ under the three correlation structures as indicated by the three points listed above. This constitutes the next section.

$E(\hat{P})$ AND $\text{Var}(\hat{P})$

Define $V = \sum_{i=1}^{m} \Delta_i$. Then,

$$E(\hat{P}) = E(\hat{P} \mid V \geq 1)\text{pr}\{V \geq 1\} + \text{pr}\{V = 0\}.$$

Now,

$$\text{pr}\{V = 0\} = (1 - \pi)^m.$$

Furthermore,

$$E(\hat{P} \mid V \geq 1) = E_V \, E_{(\Delta, U)}\left[\sum_{i=1}^{m} \frac{U_i \Delta_i}{V} \mid V \geq 1\right] = E_V \, E_U \left[\sum_{\delta=1}^{V} \frac{U_{i_\delta}}{V} \mid V \geq 1\right]$$

where $(U_{i_1}, \ldots, U_{i_v})$ is the vector of the u-values obtained by the subject on his or her v-visits, $v \geq 1$. Thus, $E(\hat{P} \mid V \geq 1) = p$ and

$$E(\hat{P}) = p\left[1 - (1 - \pi)^m\right] + (1 - \pi)^m = p + q(1 - \pi)^m. \tag{5}$$

Derivation of $\text{Var}(\hat{P})$ is considerably more complicated. Thus, some preliminary considerations are in order prior to obtaining $\text{Var}(\hat{P})$ for the three cases that are dealt with in this chapter. Let the random vector $\underline{U}'$ be a single m dimensional observation from a population whose cd $\overline{f}$ is $F_{\underline{U}'}(\underline{U}')$, that is, $\underline{U}'$ is multivariate in nature. Suppose, $E(U_i^2) < \infty$, $i = 1, \ldots, m$ and let $\rho_{ij} = \text{corr}(U_i, U_j)$, $i \neq j$ and suppose, in total generality, $E(Ui)$ = µi and Var $U_i = \sigma_i^2$ i =1, ..., m. Assume, further, a hypergeometric sampling process such that $(U_{\kappa_1}, \ldots, U_{\kappa_v})$ are sampled from $(U_1, \ldots, U_m)$, $v \leq m$ without replacement. Then if $S = \sum_{j=1}^{v} U_{\kappa_j}$ we may rewrite S as $S = \sum_{i=1}^{m} U_i I(U_i)$ where

$$I(U_i) = \begin{cases} 1 & \text{if } U_i \text{ is selected in the sample} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The sampling process is assumed independent of $U_1, \ldots, Um$. That is,

$$E\left(U_i I(U_i)\right) = E(U_i) E(I(U_i)). \quad \text{Note: } E(I(U_i)) = \text{pr}\{I(U_i) = 1\} = v/m.$$

85

Thus,

$$E(U_i I(U_i)) = \mu_i \frac{v}{m}$$

Hence,

$$E(S) \quad = \quad E\left(\sum_{i=1}^{m} U_i I(U_i)\right) = \frac{v}{m}\sum_{i=1}^{m}\mu_i \tag{7}$$

$$E(S^2) \quad = \quad E\left(\sum_{i=1}^{m} U_i I(U_i)\right)^2$$

Now,

$$\left(\sum_{i=1}^{m} U_i I(U_i)\right)^2 = \sum_{i=1}^{m} U_i^2 I^2(U_i) + 2\sum_{i<j} U_i U_j I(U_i)I(U_j)$$

Furthermore, $I^2(U_i) = I(U_i)$ since $I(U_i)$ takes only the values 0 and unity; and $I(U_i)I(U_j) = 1$ if both U_i and U_j are selected in the sample. Otherwise, $I(U_i)I(U_j) = 0$. We note,

$$E(I(U_i)I(U_j)) = \mathrm{pr}\{I(U_i) = I(U_j) = 1\} = \left(\frac{v}{m}\right)\left(\frac{v-1}{m-1}\right).$$

Finally, we obtain

$$E(S^2) \quad = \quad \sum_{i=1}^{m} E(U_i^2)E(I(U_i)) + 2\sum_{i<j} D(U_i U_j)E(I(U_i))E(I(U_j))$$

$$= \quad \frac{v}{m}\sum_{i=1}^{m} E(U_i^2) + 2\frac{v}{m}\left(\frac{v-1}{m-1}\right)\sum_{i<j} E(U_i U_j) \tag{8}$$

$$\mathrm{Var}\, S \quad = \quad \frac{v}{m}\left[\sum_{i=1}^{m} E(U_i^2) + 2\left(\frac{v-1}{m-1}\right)\sum_{i<j}(U_i U_j) - \frac{v}{m}\left(\sum_{i=1}^{m}\mu_i\right)^2\right]$$

Consider, again, the estimator for interest, i.e.,

$$\hat{P} = \begin{cases} \sum_{i=1}^{m} U_i \Delta_i / V, & \text{if } V \geq 1, \\ 1, & \text{if } V = 0 \end{cases} \tag{4'}$$

Recall, the three correlation structures among the U_i's of interest, i.e., (1), (2), and (3).

86

In the first case, $E(U_i) = p$, $\mathrm{Var}\, U_i = pq$ and $\mathrm{corr}(U_i,\ U_j) = \rho^{|i-j|}$, $i \neq j$, $0 \leq \rho < 1$. A useful result that is easily verified is

$$\sum_{i<j} \rho^{|i-j|} = \frac{\rho}{(1-\rho)} \left[m - \left(\frac{1-\rho^m}{1-\rho} \right) \right]. \tag{9}$$

An application of (8) and (9) then yields

$$\mathrm{Var}\left(\sum_{i=1}^{m} U_i \Delta_i \mid v \right) = pq \left\{ v + \frac{2v(v-1)}{m(m-1)} \left(\frac{\rho}{1-\rho} \right) \left[m - \left(\frac{1-\rho^m}{1-\rho} \right) \right] \right\}. \tag{10}$$

Thus,

$$\mathrm{Var}(\hat{P} \mid v) = \frac{pq}{v} \left\{ 1 + \frac{2(v-1)}{m(m-1)} \right\} \left(\frac{\rho}{1-\rho} \right) \left[m - \left(\frac{1-\rho^m}{1-\rho} \right) \right] \tag{11}$$

Unconditioning on V except requiring $v \geq 1$ yields

$$\mathrm{Var}(\hat{P} \mid V \geq 1) = \frac{2pq}{m(m-1)} \left(\frac{\rho}{1-\rho} \right) \left\{ m - \left(\frac{1-\rho^m}{1-\rho} \right) \right\} + \tag{12}$$
$$+ \ pq \left[1 - \frac{2}{m(m-1)} \left(\frac{\rho}{1-\rho} \right) \left\{ m - \left(\frac{1-\rho^m}{1-\rho} \right) \right\} \right] E(V^{-1}).$$

In the second case, $E(Ui) = p$, $\mathrm{Var}\ Ui = pq$ and $corr(U_i, U_j) = \rho$, $i \neq j$, $0 \leq \rho < 1$. Thus, one can show, without difficulty,

$$\mathrm{Var}\left(\sum_{i=1}^{m} U_i \Delta_i \mid v \right) = vpq[1 + (v-1)\rho]. \tag{13}$$

Hence,

$$\tag{14}$$

Finally, in the third case, $E(U_i) = p$, $\mathrm{Var}(U_i) = pq$ and $\mathrm{corr}(U_i, U_{i+1}) = \rho$, $0 \leq \rho < 1$, $i = 1, \ldots, m-1$ and $corr(U_i, U_j) = 0$, $j > i + 1$, $i = 1, \ldots, m-2$. Here,

$$\tag{15}$$

Define, generically, $\theta = \mathrm{Var}(\hat{P} \mid v \geq 1)$ to represent $\mathrm{Var}(\hat{P} \mid v \geq 1)$ for all three cases of interest. It then follows that, unconditionally,

$$\mathrm{Var}(\hat{P}) = \left[\theta + q^2(1-\pi)^m \right] \left[1 - (1-\pi)^m \right]. \tag{16}$$

Thus, to review, $E(\hat{P})$ is given by (5) and $\text{Var}(\hat{P})$ is given by (16).

It can be shown without much difficulty that if θ is given (12) or (15) $\lim_{n \to \infty} \text{Var}(\hat{P}) = 0$ implying that $\hat{P}$ is a consistent estimator of p assuming $\pi < 1$. On the other hand, if θ is given by (14), then consistency does not hold.

Finally, if θ is given by (12) it is noted that at $p = 0$,

$$\theta = pq E(V^{-1}) \qquad (17)$$

and at $\rho = 1$,

$$\theta = pq. \qquad (18)$$

If

$$f(\rho) = \left(\frac{\rho}{1 - \rho} \right) \left\{ m - \left(\frac{1 - \rho^m}{1 - \rho} \right) \right\},$$

it can be demonstrated with some difficulty that f(p) is an increasing function of ρ, $0 \le \rho \le 1$ and so the bounds on $\text{Var}(\hat{P})$ are

$$\left[pq E(V^{-1}) + q^2 (1 - \pi)^m \right] \left[1 - (1 - \pi)^m \right] \le \text{Var}(\hat{P})$$
$$\le \left[pq + q^2 (1 - \pi)^m \right] \left[1 - (1 - \pi)^m \right]. \qquad (19)$$

Consideration of $f(\rho)$ is contained in the appendix. Finally, Mendenhall and Lehman (1960) show that

$$E(V^{-1}) \doteq \left(\frac{m - 2}{m} \right) (m\pi - (1 - \pi))^{-1} \qquad (20)$$

and provides two significant figure accuracy for $m\pi > 5$.

COMPARISON OF TREATMENT GROUPS

The goal of this section is to develop a test of the hypothesis H_0: $p_1 = p_2$, where p_i is the probability an individual in the ith treatment group tests positively for the presence of the opiate. It is noted, in the example to be presented in the next section, that there are three treatment groups in question; therefore, multiple comparison methods are used in comparing the results among the three groups. In this section, the notation adds subscripts κj to represent the jth time the κth individual is tested within a treatment group.

Let $\hat{P}_{i\kappa}$ be the proportion of the trials in which the κth individual tests positive on the ith treatment, $\kappa = 1, \ldots, n_i$. Then, it is clear from (4) that

$$\hat{P}_{i\kappa} = \begin{cases} \sum_{j=1}^{m_{i\kappa}} \frac{U_{i\kappa j} \Delta_{i\kappa j}}{V_{i\kappa}}, & \text{if } V_{i\kappa} \ge 1 \\ 1, & \text{if } V_{i\kappa} = 0. \end{cases} \qquad (4')$$

88

Define $\hat{P}_i = \sum_{\kappa=1}^{n_i} \hat{P}_{i\kappa}/n_i$. Again, note that

$$(5')$$

since within each treatment group it is assumed $p_{il} = \cdots = p_{im_{i\kappa}} = p_i$. Thus,

$$E(\hat{P}_i) = p_i + q_i \sum_{\kappa=1}^{n_i} \frac{(1 - \pi_i)^{m_{i\kappa}}}{n_i}. \tag{19}$$

If the same reasoning is followed concerning $\mathrm{Var}\hat{P}_{i\kappa}$, one finds

$$\mathrm{Var}\, \hat{P}_{i\kappa} = \left[\theta_{i\kappa} + q_i^2(1 - \pi_i)^{m_{i\kappa}}\right]\left[1 - (1 - \pi_i)^{m_{i\kappa}}\right] \tag{16'}$$

where, generically, $\theta_{i\kappa} = \mathrm{Var}(\hat{P}_{i\kappa} \mid V_{i\kappa} \geq 1)$. Since $\hat{P}_{i\kappa}$ and $\hat{P}_{i\kappa'}$ are stochastically independent for $\kappa \neq \kappa'$, it follows that

$$\mathrm{Var}\, \hat{P}_i = \sum_{\kappa=1}^{n_i} \frac{\left[\theta_{i\kappa} + q_i^2(1 - \pi_i)\right]^{m_{iv}} \left[1 - (1 - \pi_i)\right]^{m_{i\kappa}}}{n_i^2}. \tag{20}$$

It is easy to show that for all three cases, i.e., all three $\theta_{i\kappa}$, $\mathrm{Var}\, \hat{P}_i \to 0$ as $n_i \to \infty$, $i = 1, 2, 3$. Thus, $\hat{p}_i$ is a consistent estimator of $p_i + q_i a_i$ where

$$a_i = \lim_{n_i \to \infty} \sum_{\kappa=1}^{n_i} \frac{(l - \pi_i)^{m_{i\kappa}}}{n_i}.$$

In order to reduce the bias, let

$$\hat{P}_i^* = \hat{P}_i - \hat{Q}_i \sum_{\kappa=1}^{n_i} \frac{(1 - \hat{\pi}_i)^{m_{i\kappa}}}{n_i}. \tag{21}$$

$\hat{P}_i^*$ is termed the reduced estimator and its realizations are presented as reduced estimates in table 2, $i = 1, 2, 3$.

The variance of (21) can be approximated by the delta method. Thus,

$$\hat{\mathrm{Var}}\hat{P}_i^* = \hat{\mathrm{Var}}\hat{P}_i \left\{1 + \sum_{\kappa=1}^{n_i} \frac{(1 - \hat{\pi}_i)^{m_{iv}}}{n_i}\right\}^2$$

$$+ \hat{\mathrm{Var}}\hat{\pi}_i \hat{q}_i^2 \left\{\sum_{\kappa=1}^{n_i} \frac{m_{i\kappa}(1 - \pi_i)^{m_{i\kappa}-1}}{n_i}\right\}^2. \tag{22}$$

Hence, to test for treatment difference, the following confidence intervals all at $(1 - \alpha/6)$ can be used:

$$(\hat{p}_i^* - \hat{p}_j^*) \pm Z_{\alpha/6}\sqrt{\hat{\mathrm{Var}}\hat{P}_i^* + \hat{\mathrm{Var}}\hat{P}_j^*} \tag{23}$$

$i \neq j; i, j = 1, 2, 3,$ where $Z_{\alpha/6}$ is the $\alpha_l/6$th percentile of the standard normal pdf. Thus, confidence intervals for $p_1 - p_2$, $p_1 - p_3$, and $p_2 - p_3$ can be constructed using (23) with overall confidence of at least $(1 - \alpha)$.

ANALYSIS OF THE EXISTING DATA AND CONCLUSIONS

The methodology that has been developed in this chapter is now applied to the double blind, three-armed controlled ARC trial. This trial was conducted to evaluate the efficacy of buprenorphine (arm 1), methadone 20 mg (arm 2), and methadone 60 mg (arm 3) in the treatment of opiate addiction. Data on only the first 17 weeks of the study were used to study how well the patients were maintained on the treatment drug. No analysis was performed on weeks 18 through 27.

Table 1 shows the summary results for the three treatment groups.

The correlation coefficient in each treatment group was estimated as an average (unweighted) of the serial correlations for the patients in that group. It should be noted from table 1 that there is no statistically significant difference among the π''s. That is, roughly 77 percent of all individuals presented urine samples in the first 17 weeks of the study. Furthermore, it can also be shown that all three of the correlation coefficients are not statistically significantly different from zero. However, it was decided to use the established values of ρ for illustrative purposes.

TABLE 1. *Estimates of p, π, and ρ*

Treatment Group	$\hat{p}$	$\hat{\pi}$	$\hat{\rho}$	n
Buprenorphine	0.483	0.773	0.087	53
Methadone 20 mg	0.687	0.767	0.013	55
Methadone 60 mg	0.564	0.786	0.133	54

TABLE 2. *Estimates of the probability of opiate use*

Treatment Group	$\hat{p}$	$\text{Var}_1\hat{P}$	$\text{Var}_2\hat{P}$	$\text{Var}_3\hat{P}$
Buprenorphine (raw)	0.483	$3.83 \cdot 10^{-4}$	$6.76 \cdot 10^{-4}$	$3.33 \cdot 10^{-4}$
Methadone 20 mg (raw)	0.687	$3.80 \cdot 10^{-4}$	$4.15 \cdot 10^{-4}$	$3.51 \cdot 10^{-4}$
Methadone 60 mg (raw)	0.564	$4.79 \cdot 10^{-4}$	$9.05 \cdot 10^{-4}$	$4.16 \cdot 10^{-4}$
Buprenorphine (reduced)	0.468	$4.05 \cdot 10^{-4}$	$7.08 \cdot 10^{-4}$	$3.52 \cdot 10^{-4}$
Methadone 20 mg (reduced)	0.685	$3.87 \cdot 10^{-4}$	$4.22 \cdot 10^{-4}$	$3.57 \cdot 10^{-4}$
Methadone 60 mg (reduced)	0.558	$4.90 \cdot 10^{-4}$	$9.25 \cdot 10^{-4}$	$4.26 \cdot 10^{-4}$

Table 2 provides the raw and the reduced estimates of p_i, $i := 1, 2, 3$, the probability of detecting the opiate in each of the three treatment groups. Also, the three estimates of variance are provided for the three different patterns of correlation that are assumed.

Where $\text{Var}_1\hat{P}$ assumes visits i and j have correlation $\rho^{|i-j|}$, $\text{Var}_2\hat{P}$ assumes all visits, adjacent as well as nonadjacent, have correlation ρ and, finally, $\text{Var}_3\hat{P}$ is such that adjacent visits have correlation ρ and nonadjacent visits have correlation zero.

Overall, 95 percent confidence intervals for $p_l - p_\kappa$, $\kappa \neq l$; $\kappa, l = 1, 2, 3$ are then easily obtained for each pair of treatment differences. The formula used is

$$(\hat{p}_l - \hat{p}_\kappa) \pm Z_{0.0083}\sqrt{\hat{\text{Var}}_j\hat{p}_l + \hat{\text{Var}}_j\hat{p}_\kappa},$$

to ensure an overall 95 percent confidence, $j = 1, 2, 3$, the three variance estimates based on the three correlation patterns assumed. Note that $Z_{0.00833} = 2.409$.

As one would expect, the largest variance occurs when ρ is constant across all visits, has a value between the smallest and largest values when visits i and j have correlation $\rho^{|i-j|}$, and is smallest when adjacent visits have correlation ρ and nonadjacent visits have zero correlation.

If one examines the confidence intervals that are generated from table 2 (see table 3) it is clear that buprenorphine is superior to methadone 20 mg, regardless of the correlation pattern assumed. Furthermore, methadone 60 mg is clearly superior to methadone 20 mg. However, it is still not clear that buprenorphine is a better treatment regime than methadone 60 mg. However, the analysis is suggestive of this conclusion, since the only situation where the confidence interval contains the null hypothesis is when the correlation between

TABLE 3. *Confidence intervals of the difference of the probabilities of opiate use*

Group Difference	$c.i_1.$	$c.i_2.$	$c.i_3.$
Meth 20 - bup (raw)	0.137-0.271	0.124-0.284	0.141-0.267
Meth 20 - meth 60 (raw)	0.052-0.194	0.036-1.210	0.056-0.190
Meth 60 - bup (raw)	0.010-0.152	-0.014-0.176	0.015-0.147
Meth 20 - bup (reduced)	0.149-0.285	0.136-0.298	0.153-0.281
Meth 20 - meth 60 (reduced)	0.056-0.198	0.039-0.215	0.060-0.194
Meth 60 - bup (reduced)	0.018-0.162	-0.007-0.187	0.023-0.157

pairs of visits remains constant over all pairs of visits. If, however, this is the situation, then mathematically random testing vs. systematic testing does not make a great deal of difference, since fixed as well as randomly spaced times between visits will have the same correlation.

It should be noted at this point that the true correlational structure between pairs of visits does not and should solely determine whether random or systematic testing is appropriate. If the sampling scheme is determined on the basis that an individual is still using a given substance during a given week or throughout the study period, then random sampling is likely appropriate to detect his or her use. However, if the extent of drug abuse is of importance, such as with the ARC 090 study, then capturing all the episodes of drug abuse is important and, hence, systematic sampling is likely to be more useful since drug-seeking behavior is *not* random. Finally, it is noted that carryover effects are probably quite important but are not considered here.

APPENDIX

Theorem: Suppose

$$f(\rho) = \frac{\rho}{1-\rho} m - \left[\frac{1-\rho^m}{1-\rho} \right]. \qquad (A.1)$$

Then $f(0) = 0$ and $f(1) = \frac{m(m-1)}{2}$ and $f(\rho)$ increases in ρ, $0 < \rho < 1$.

Proof: $f(0) = 0$ is trivial. To find $f(1)$ an application of L' Hôpital's rule twice is needed. Finally, to show $f(\rho)$ increases in ρ, it is noted that

$$\ln f(\rho) = \ln \rho + 2 \ln(1-\rho) + \ln\left[(m-1) + \rho^m - \rho\right]. \qquad (A.2)$$

Thus,

$$\frac{d\ln f(\rho)}{d\rho} = \frac{(m-1)[1-\rho^{m+1}] - (m+1)\rho[1-\rho^{m-1}]}{\rho(1-\rho)\{m-1+p^m - m\rho\}}.$$

The denominator is positive for $0 < \rho < 1$ since $m(1 - p) > 1 - \rho'^m$ for ρ in this domain. This is easily established by induction on m.

Finally, it is necessary to establish that

$$(m-1)[1-\rho^{m+1}] - (m+1)\rho[1-\rho^{m-1}] \geq 0 \qquad (A.3)$$

for $m > 1$ and $0 \leq \rho \leq 1$. If $g(\rho)$ is defined as

$$g(\rho) = (m-1) + (m+1)\rho^m - (m-1)\rho^{m+1} - (m+1)\rho, \qquad (A.4)$$

then $g(0) = m - 1$ and $g(1) = 0$. To complete the establishment of (A.3), it suffices to show $g(\rho)$ is monotone on $0 \leq \rho \leq 1$. To this end,

$$g'(\rho) = (m+1)[m\rho^{m-1}(1-\rho) + \rho^m - 1] \leq 0$$

for $0 \leq \rho \leq 1$, which completes the demonstration.

ACKNOWLEDGMENTS

REFERENCES

Council on Science Affairs, Scientific issues in drug testing. *JAMA* 257(22):3110-3114, 1987

Gabriel, K.R. The distribution of the number of successes in a sequence of dependent trials. *Biometrika* 46:454-460, 1959.

Gabriel, K.R., and Neuman, J. A Markov chain model for daily rain-fall occurrence at Tel-Aviv. *Q J Roy Meteorol Soc* 88:90-95, 1962.

Haseman, J.K., and Kupper, L.L. Analysis of dichotomous response data from certain toxicological experiments. *Biometrics* 35(1):281-293, 1979.

Klotz, J. Statistical inference in Bernoulli trials with dependence. *Ann Stat* 2(1):373-379, 1973.

Kupper, L.L., and Haseman, J.K. The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* 34(1):69-76, 1978.

Mendenhall, W., and Lehman, E.H. An approximation to the negative moments of the positive binomial useful in life testing. *Technometrics* 2:233-239, 1960.

Rosner, B. Multivariate methods in ophthalmology with application to other paired-data situations. *Biometrics* 40(4):1025-1035, 1984.

AUTHOR

Alan J. Gross, Ph.D.
Professor
Department of Biostatistics, Epidemiology, and Systems Science
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2503

Summary of Discussion: "Three Estimators of the Probability of Opiate Use From Incomplete Data" by Gross

Ram B. Jain

Dr. Redmond, who reviewed Dr. Gross' chapter prior to the technical review meeting, questioned the appropriateness of the assumption that data are missing at random. Dr. Gross agreed that this assumption may not be appropriate and that further work needs to be done in this area.

Dr. Fisher warned against jumping to conclusions too soon even if the missing value and/or dropout rates are the same across different treatment groups. In the placebo group, for example, patients may miss visits and/or drop out because of the lack of efficacy, whereas in the other groups, they may miss visits and/or drop out because of adverse events. These different reasons for the *same* missed visits and/or dropout rates must have a bearing on the inference that is drawn and must be carefully looked into since they *may* have different implications for different treatments.

Dr. Redmond also raised the issue of using an estimator in which one looks at the average of visits across time. Two treatments might result in the same average across visits over time, but for example, positive urines may be clustered at the end of the visits in one case and toward the beginning or scattered throughout in the other case. An estimator that looks only at the average across visits over time would not be able to discriminate this different pattern of positive urines over time in the two treatments.

I suggested that a model be considered that incorporates possible different correlational structures for different segments of the study. The addicts do not suddenly stop abusing the drugs, because drug abuse medications take time to work. During the first segment of the study, there probably will be longer sequences of positive urines with *occasional* negative urines. This will be followed by a probably random pattern of negative and positive urines indicating the medication has started working. During the last segment of the study, if the medication did work, there probably will be longer sequences of

95

negative urines with occasional positive urines. Dr. Gross indicated that such a model may be possible if a close form variance estimator can be obtained; for example, different correlational structures suggested by him in his chapter can be used for different Segments Of the study or a piece-wise fitting Of the data can be attempted (see Weng, this volume).

Dr. Hedayat suggested that the patient characteristics, for example, sex or age, be included in the model rather than reduce the problem to only a few factors, that is, p's, π's, ¡and p's as in Dr. Gross' chapter. One of the participants also suggested that visit number be used as a covariate since the probabilities of a positive urine are likely to change over time.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Issues in the Analysis of Clinical Trials for Opiate Dependence

Dean Foilmann, Margaret Wu, and Nancy Geller

INTRODUCTION

The analysis of clinical trials of opiate dependence presents special challenges. In particular, the amount of missing data tends to be substantial. In trials of cardiovascular disease, dropout rates higher than 5 percent per year are considered high. In contrast, in a recent trial of three treatments for heroin addicts, ARC 090, more than two-thirds of the patients dropped out by the end of 17 weeks. In such a trial, any pronouncement of treatment efficacy depends on how one deals with missing data.

There are two main possibilities for dealing with missing endpoint data in a clinical trial: ignoring them or using imputation or modeling. Ignoring the missing data can lead to bias in inference if the data are not missing at random and can violate the intent-to-treat principle, which demands that all subjects be analyzed as part of the treatment group to which they were randomly assigned. Imputation or modeling is inherently assumption dependent but can provide accurate answers if the assumptions are met. A problematic feature of opiate-dependence trials is that some assumptions are completely uncheckable. In essence, a guess is made about drug use on days of missed visits. Although the guess can be based on other data or informed opinion of observed behavior, the degree to which the guess reflects unobserved behavior is unknowable.

Wu and Carroll (1988), Wu and Bailey (1989) and Wu and colleagues (1991) considered the effect of censoring when comparing changes Of a continuous response variable between treatment groups. They used the term "informative censoring" when the probability of a missing data point depends on the parameter of interest. They showed that, when the censoring is informative, falsely assuming random censoring could give biased estimates of the group slope means and the between-group differences. Statistical procedures to account for informative censoring were also proposed.

This chapter discusses different methods of dealing with longitudinal data with many values missing. It is assumed that individuals are supposed to be tested at a fixed number of times (e.g., Monday, Wednesday, and Friday of each week for 17 weeks) and that each test is either positive, indicating drug use, or negative, indicating nonuse. However, because some individuals do not show up for some tests, the test value for that visit is missing. For simplicity, the authors' primary method of analysis will compute a summary measure of each individual's sequence of tests and compare these measures between the two groups.

A second section of this chapter considers different methods of imputation and modeling for these data, which are considered to be binary repeated measurements. Notation is developed and the case of no missing data discussed. Ad hoc tests for missing data are explored, followed by the simple approach of imputing a specific value to all missing observations. Combining test statistics for treatment efficacy and equal missing data is considered. Next, models for informative censoring are introduced whereby individuals' imputed values depend on their observed data. This generalizes previous work to the case of binary data. Section three applies some of these methods to the data of ARC 090. The discussion section suggests some tests for more complicated null hypotheses.

NOTATION

The authors assume that the subjects are randomly assigned to one of two treatment groups with subjects in group k = 1,2. The data of interest consist of the repeated binary measurements where

$$Y_{ijk} = \left\{ \begin{array}{ll} 1 & \textit{if the jth test of the ith subject in group k is positive} \\ 0 & \textit{otherwise} \end{array} \right.$$

for j = 1,. . .,m, and i = 1,. . .,n_k.

In the background, we imagine a population of $(\theta_{i1k}, \ldots, \theta_{imk})$ vectors that govern the response of the ith subject in the kth group. These vectors are paired with subjects via the randomization process. For notational convenience, we define $P(Y_{ijk} = 1) = \theta_{ijk}$.

Because subjects can miss visits as well as drop out of the study, define

$$D_{ijk} = \left\{ \begin{array}{ll} 1 & \textit{if the jth test of the ith subject in group k is absent} \\ 0 & \textit{otherwise} \end{array} \right.$$

Similar to the θ_{ijk} notation, let $P\{D_{ijk} = 1\} = \pi_{ijk}$. We also define, for the ith subject in the kth group, L_{ik}: the last visit number and m_{jk} as the number of available test results.

For simplicity, it is assumed that the hypothesis of interest is whether the average θ_{ij} differs between the two groups, A useful notation for this is $H_0^{(1)}: E[\theta_{ij1}] = E[\theta_{ij2}]$, where the expectation is over the subscripts (here i and j). Note that each subject is allowed a different average response proportion and response proportions are averaged over time. The probability of a positive response within a subject over time may or may not be the same. With no missing data, a reasonable test of $H_0^{(1)}$ is the standardized difference in the mean proportion of positive tests of the two samples

$$T_1 = \frac{\sum_{i=1}^{n_1} \bar{Y}_{i \cdot 1}/n_1 - \sum_{i=1}^{n_2} \bar{Y}_{i \cdot 2}/n_2}{\hat{\sigma}_{T_1}} \tag{1}$$

where $\bar{Y}_{i \cdot k}$ is the proportion of positive tests for the ith subject in the kth group, and

$$\hat{\sigma}_{T_1}^2 = \sum_{i=1}^{n_1}(\bar{Y}_{i \cdot 1} - \bar{Y}_{\cdot \cdot 1})^2/(n_1 - 1) + \sum_{i=1}^{n_2}(\bar{Y}_{i \cdot 2} - \bar{Y}_{\cdot \cdot 2})^2/(n_2 - 1) \tag{2}$$

where $\bar{Y}_{\cdot \cdot k}$ is the average of the $\bar{Y}_{i \cdot k}$s.

Alternately, $\hat{\sigma}_{T_1}^2$ may be taken as the pooled sample estimate of the variance. With either estimate of variance, T_1 has an asymptotic standard normal distribution. Another reasonable test in this situation is the two-sample Wilcoxon test.

AD HOC TESTS WITH INCOMPLETE DATA

If some of the tests are missing, (1) may require modification. Suppose that $\theta_{ijk} = \theta_{ik}$ (i.e., that there is no variation in the probability of a positive test over time), but again different subjects are allowed different positive propensities. Implicit in this formulation is that the observable data on an individual reflect the unobservable data on that individual. Note, however, that these assumptions do allow for a dependency between the probability of a missing test and the probability of a positive response. An unbiased estimate of an individual's θ_{ik} here is the average response on completed visits or $\bar{Y}_{i \cdot k} = \sum_j Y_{ijk}/m_{ik}$. A simple way to compare these proportions in the two groups is to use (1) with a modified variance estimate. In effect, the authors are comparing the unweighted within-group averages $\bar{Y}_{\cdot \cdot k} = \sum_i \bar{Y}_{i \cdot k}/n_k$ and testing a new hypothesis $H_0^{(2)}: E[\theta_{i1}] = E[\theta_{i2}]$.

A consistent variance estimate can be derived under the assumption $H_0^{(2)}$. Since the θ_{ik} are allowed to vary within each group, a random effects model is specified for the θ_{ik} for each k. The mean and variance of θ_{ik} are denoted by μ_k and τ_k^2. The following expectations can be used to provide simple method-of-moments estimates of μ_k and τ_k^2:

$$E\ [\bar{Y}_{i \cdot k}] = \mu_k$$

$$E\ [\sum_i (\bar{Y}_{i \cdot k} - \bar{Y}_{\cdot \cdot k})^2 / (n_k - 1)] = W_k(\mu_k(1 - \mu_k) - \tau_k^2) + \tau_k^2$$

$$E\ [\sum_{ij}(Y_{ijk} - \bar{Y}_{i \cdot k})^2 / \sum_i (m_{ik} - 1)] = \mu_k(1 - \mu_k) - \tau_k^2$$

where $W_k = 1/n_k \sum_i 1/m_{ik}$. The last two equations can be used to estimate τ_k^2.

Under the random effects model,

$$V\ [\bar{Y}_{i \cdot k}] = (1/m_{ik})\ [\mu_k(1 - \mu_k) + (m_{ik} - 1)\tau_k^2],\ \ (3)$$

and an estimate of $V[\bar{Y}_i]$, say, $\hat{V}[\bar{Y}_i]$ can be gotten by estimating μ_k and τ_k^2. With missing data, the denominator of (1) is then estimated by the square root of $\hat{\sigma}_{T_2}^2 = \sum_i \hat{V}\ [\bar{Y}_{i \cdot 1}] + \sum_i \hat{V}\ [\bar{Y}_{i \cdot 2}]$. Call this test T_2.

One variation on this theme is to use a weighted comparison of the average responses where $\bar{Y}_{i \cdot k}$ is weighted by m_{ik}. However, this weighted analysis requires the stronger assumption that the probability of missing a visit or dropping out is unrelated to a person's average response. If persons who drop out early have a higher proportion of negative tests, the weighted estimate will tend to be biased downward. However, if the additional assumption is warranted, the weighted analysis can be more efficient than that based on the unweighted average.

Another method is to calculate ranks based on the $\bar{Y}_{i \cdot k}$ and to perform a rank test. This approach is appealing since it is based directly on the unbiased response averages for each subject, Furthermore, ranks tend to dampen the influence of the more variable $\bar{Y}_{i \cdot k}$ based on few observations (Wu et al. 1991). Suppose the proportion positive is ranked from lowest to highest. Tied observations can be resolved by evaluating the number of complete

observations and whether the ties are above or below the overall mean $\bar{Y} = \sum_{k=1}^{2}\sum_{i=1}^{n_k} \bar{Y}_{i\cdot k}/(n_1+n_2)$. If $\bar{Y}_{i\cdot k} = \bar{Y}_{\iota\cdot k} > \bar{Y}$, and $m_{ik} > m_{\iota k}$, take $Rank\,(\bar{Y}_{i\cdot k}) <$ $Rank\,(\bar{Y}_{\iota k})$ If the equal sample averages are below the overall mean, take $Rank\,(\bar{Y}_{\iota\cdot k}) > Rank\,(\bar{Y}_{i\cdot k})$. Treating ties in this manner is consistent with the ranking based on the shrinkage or Empirical Bayes estimates of the θ_is. This generalizes a method for breaking ties suggested by Sahlroot and Pledger (1991). Call this rank test T_3.

The assumption that $\theta_{ijk} = \theta_{ik}$ may be untenable. For example, the chance of a positive test may increase as the study progresses and subjects lose interest. Even in this case, the subject average $\bar{Y}_{i\cdot k}$ provides an unbiased estimate of his or her average θ_{ijk} over the complete visits. Therefore, tests of the hypothesis that

$$H_0^{(3)}: E\,[\theta_{ij1}(1-D_{ij1}) \mid D_{ij1}] = E\,[\theta_{ij2}(1-D_{ij2}) \mid D_{ij2}]$$

can in principle be made by looking at either the ranks of the $\bar{Y}_{i\cdot k}$s or the difference of the unweighted within-group averages. One problem with the former test is that a variance estimate requires additional structure to be put on the θ_{ijk}s. Although the test based on ranks can be used in a straightforward manner, $H_0^{(3)}$ may be of limited interest if the distribution of the D_{ij1}s differs from that of the D_{ij2}s. For example, if θ_{ijk} increases with j for both groups and subjects drop out early for one group, the group with the earlier dropouts will tend to have a lower test average. The test statistic could identify the better treatment as the one with the earlier dropouts even if later θ_{ijk}s were higher for the group with earlier dropouts.

Since the rank test of $H_0^{(3)}$ may result in a misleading inference, it is necessary to examine whether the missed visits and dropout times differ between the two groups. Tests of equality of the pattern of missingness between the two groups should be calculated. Rank tests, tests of means, or logrank tests could be used. Informally, these tests could be used to see how meaningful the test of $H_0^{(3)}$ is. For example, if a similar pattern of missingness is expected to occur over time in the treatment groups, one could use a test of the difference in proportions of missing data in the two treatment groups, which is analogous to (1) with $\bar{D}_{i\cdot k}$ replacing $\bar{Y}_{i\cdot k}$ and a suitable estimates of variance, say $\hat{\sigma}_{M_1}$ replacing $\hat{\sigma}_{T_1}$. Call this test statistic M_1. A Wilcoxon test based on $\bar{D}_{i\cdot k}$ is also considered; call this test statistic M_2.

More formally, a test for missingness can be combined with a test of efficacy using a multivariate test (O'Brien 1984; Pocock et al. 1987). Here the null hypothesis is

$$H^{(4)}:(E\ [\theta_{i1}], E\ [\pi_{ij1}]) = (E\ [\theta_{i2}], E\ [\pi_{ij2}])$$

and the alternative is

$$H_A^{(4)}:(E\ [\theta_{i1}], E\ [\pi_{ij1}]) > (E\ [\theta_{i2}], E\ [\pi_{ij2}])\ \text{or}$$

$$(E\ [\theta_{i1}], E\ [\pi_{ij1}]) < (E\ [\theta_{i2}], E\ [\pi_{ij2}]);$$

that is, one treatment is better than the other with respect to both the proportion of positive tests and the proportion of missing tests.

As an example, consider combining two rank tests, such as T_3 and M_2. O'Brien (1984) proposed ranking each outcome separately, as one would to perform a Wilcoxon rank sum test on each outcome and summing the ranks over the two outcomes, He then proposed calculating a Wilcoxon rank sum test for these sums. Call the resulting statistic 0,. In the case of more than two samples, O'Brien suggested ranking each outcome over all samples, forming the rank sums for each subject, and then using one-way analysis of variance on the sums. Alternatively, a Kruskal-Wallis test could be used on the sums.

More complex combinations of test statistics may be formed using the method proposed by Pocock and colleagues (1987). To combine T_2 with M_1 in an O'Brien-type statistic, the correlation $\rho(T_2, M_1)$ between T_2 and M_1 would need to be estimated. Pocock and colleagues (1987) gave an explicit reduction of the formula for O'Brien's generalized least-squares statistic when endpoints are equally correlated and the within-group data are iid. The within-group data here are not iid, however, since the variance of $\overline{Y}_{j \cdot k}$ depends on m_{ik}. Estimation of the correlation between these test statistics requires further work.

SIMPLE IMPUTATION

If the assumption that $\theta_{ijk} = \theta_{ik}$ is untenable, an imputation of a specific value for each missing data point may be reasonable. We will call this simple imputation. One possibility is to replace missing responses with positive responses. This is appropriate if it seems likely that subjects would have tested positive if they had

been tested. Another rationale for this imputation is that it defines a new endpoint: missed test or positive test. This endpoint tests a new hypothesis

$$H^{(5)} : E\ [\eta_{ij1}] = E\ [\eta_{ij2}],$$

where η_{ijk} is the probability that the ith person in the kth group is positive or missing for the jth test. One could argue that both positive tests as well as missing data suggest failure of the program. An advantage of the simple imputation approach is that the analysis then proceeds as if complete data had been obtained, and (1) or its rank analog can be used. The test statistic T_4 is referred to as (1) with a value 1 imputed for missing values.

MODEL-BASED IMPUTATION

The basic idea here is to use a model to provide an accurate test of the original hypothesis $H_0^{(1)} : E\ [\theta_{ij1}] = E\ [\theta_{ij2}]$ even if individuals likely to test positive tend to drop out or if $\theta_{ijk} \neq \theta_{ik}$. The authors attempt to succinctly describe θ_{ijk} for each group with a model, estimate the parameters of the model separately in each group, and then compare the estimate of $\sum_{i=1}^{n_1}\sum_{j=1}^{m}\theta_{ij1}/(n_1 m)$ to the estimate of $\sum_{i=2}^{n_2}\sum_{j=1}^{m}\theta_{ij2}/(n_2 m)$. Although this approach is heavily model based, one can allow for quite general effects of dropping out, missed visits, and other factors as long as they are correctly incorporated into the model.

To justify our procedure in a simple setting, ignore the treatment identifier k and suppose that the following model holds:

$$\ln\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right) = B + \beta_{0i} \tag{4}$$

$$\ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \alpha_0 + \alpha_1 \beta_{0i} \tag{5}$$

where β_{0i} is a random parameter with some distribution H with mean zero, and B, α_0, α_1 are fixed-effects parameters. In other words, each subject draws a random propensity for a positive test, β_{0i}, from H, which also affects the probability of a missed visit. If α_1 is not zero, the missing observations are said to be informative with respect to the parameter of interest, β_{0i} (Wu and Carroll 1988).

Although maximum likelihood estimation could be performed using the (Y_{ij}, D_{ij}), a simpler approach can be argued as was done by Wu and Bailey (1989). Suppose the Y_{ij}s are ignored. The model defined by (5) can be viewed as an Empirical Bayes model. The information from the ith individual is given by $\sum_{j=1}^{m} D_{ij}$ or equivalently m_i since $m_i = m - \sum_{j=1}^{m} D_{ij}$. For the ith individual and any prior H, the posterior expectation $E[\beta_{0i}|m_i]$ is increasing (decreasing) as a function of m_i when a, is negative (positive). A proof of this result is presented elsewhere. In other words, if α_1 is positive, individuals who are likely to test positive are also likely to miss tests.

This result suggests that a simple way to capture the information contained in is to use the model

$$\ln\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right) = B + \beta_{0i} + h(m_i)$$

where β_{0i} |has a distribution $H( )$ with mean 0, and $h( )$ is some function that is allowed to be either increasing or decreasing, such as a polynomial, perhaps with restricted coefficients.

The above argument justifies fitting a logistic regression model with random subject effects and other fixed effects that describe the missingness. Such models have been discussed in a more general setting by Pierce and Sands (1975), Stiratelli and colleagues (1984), and Follmann and Lambert (1989). The approach of Follmann and Lambert (1989) is used and H is estimated via nonparametric maximum likelihood, along with parametric estimates of the fixed effects. Under this approach, H is assumed to follow a distribution with a finite number of support points. The number of support points is estimated by the data.

For the problem at hand, consider the following model:

$$\ln\left(\frac{\theta_{ijk}}{1-\theta_{ijk}}\right) = B_k + \beta_{0ik} + \beta_{1k}m_{ik} + \beta_{2k}L_{ik} \tag{6}$$

for k = 1,2. Note that the probability of a positive response is assumed free of j and the test derived from it is appropriate for the hypothesis $H_0^{(2)}: E[\theta_{i1}] = E[\theta_{i2}]$. Therefore, results from this procedure can be compared with other tests of the same hypothesis from the previous sections.

The numerator of the model-based test statistic is

$$\sum_{i=1}^{n_1} \hat{E}\,[\theta_{i1}]/n_1 - \sum_{i=1}^{n_2} \hat{E}\,[\theta_{i2}]/n_2$$

where the expectations are Empirical Bayes posterior expectations using the within-group estimates of the random effects distribution, the fixed effects, and each individual's data. The asymptotic variance of this test statistic can be estimated by the delta method, given a covariance matrix for the estimates. The authors use the observed Fisher Information (pretending that the number of support points is known) to estimate this covariance (Follmann and Lambert 1989).

In general, trends in the θ_{ij}s could be made to depend on j via a covariate, for example, polynomials of j or $\log(j)$. If the θ_{ij}s depend on j and this dependence is accurately summarized via the random effects model, the original hypothesis $H_0^{(1)}: E\,[\theta_{ij1}] = E\,[\theta_{ij2}]$ can be tested, even with missing data.

EXAMPLE

A recent randomized clinical trial compared three treatments-buprenorphine, methadone at 20 mg (methadone 20), and methadone at 60 mg (methadone 60)-for their ability to reduce opiate use within a group of addicts. This section focuses on the buprenorphine and methadone 20 groups.

Respectively, 53 and 55 subjects were randomized to these two groups. The methadone 60 group contained 54 subjects. Following randomization, urine tests were conducted three times per week for a total of 17 weeks. One subject was assigned to the buprenorphine group who never took a test. Although one might include this subject with some imputed value in an analysis of the treatments, she is excluded for simplicity.

Figure 1 displays the proportions of positive tests over visits for the two groups. Buprenorphine is almost always better, and there seems to be little trend in the proportion positive. Figure 2 displays the proportion of missed tests over time. Both groups show increasing trends that seem fairly comparable. Figure 3 displays the scatter plot of $\bar{Y}_1$ vs. $1 - m_i/51$, that is, the proportion positive vs. the proportion missed for the two groups. A moderate positive correlation ($\rho = .51$) is seen between the two in the buprenorphine group, whereas the correlation is less strong in the methadone 20 group ($\rho = .18$). Also note that subjects in the buprenorphine group who always test positive have many missed tests. Some subjects in the methadone 20 group who always test positive rarely show up.
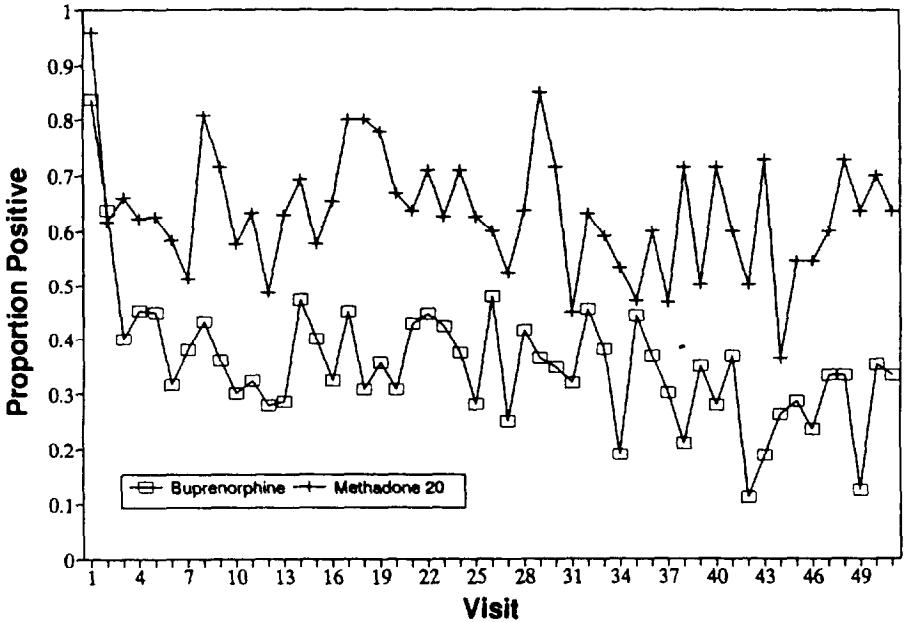
FIGURE 1. *Proportion of positive tests over time, by treatment*

Table 1 shows some sample statistics for the two groups. Burprenorphine has a lower average response proportion, a larger random-effects variance, and also a larger variance of $\sum_i \bar{Y}_{i\bullet}/n$. The latter discrepancy is influenced both by the larger $\hat{\tau}^2$ and the average response being closer to .5 for the buprenorphine group. The average proportion missing is somewhat larger in the methadone 20 group.

Table 2 provides the results for the various tests discussed in the text. For all test statistic numerators, the methadone group result is subtracted from the buprenorphine group result. The first two tests provide very similar results for the hypothesis that the average θ_i is the same for the two groups. The first test is obtained from the results of table 1. Both tests indicate that the buprenorphine group has a substantially lower probability of a positive test.

For the rank test, the authors determined the number of times that the Empirical Bayes approach adjudicated tied observations. With no ties, each of the 107 observations forms a "cluster" of 1. For these data, there were 69 clusters, ranging in size from 2 to 25. For example, for $\bar{Y}_{i\bullet} = .5$, there are four observations, one with $m_i = 4$ in the buprenorphine group and 3 in the

106
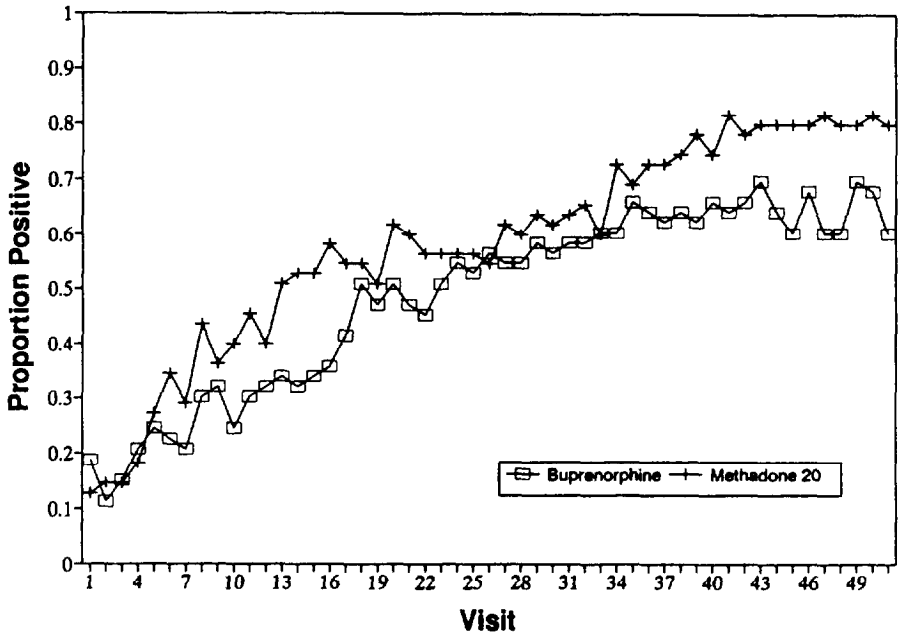
FIGURE 2. *Proportion of missed tests over time, by treatment*

methadone 20 group with m_i = 4, 2, and 2. For $\overline{Y}_{j.}$= 1.0, there were 25
observations. Following the Empirical Bayes method of breaking ties, there
were 92 clusters. Interestingly, when the ties are not broken, the Wilcoxon rank
test is -2.85. Thus, how ties are treated makes a difference here.

Although the overall proportion missing is somewhat higher in the methadone
group, a test of the difference in these proportions is not significant. However,
this difference explains why the test statistic that imputes 1 for missing
observations is higher than the analogous test without imputation.
Burprenorphine is better both with respect to missing data and with respect to
the proportion of positive tests.

The O'Brien rank test shows that buprenorphine was better than methadone 20
simultaneously with respect to efficacy and missingness. In calculating this
statistic, average ranks were used for ties. The final ranking had 80 clusters, 1
of size five, 4 of size three, and 15 of size two. It is not surprising that the value
of this test statistic is smaller in absolute value than the rank test for the
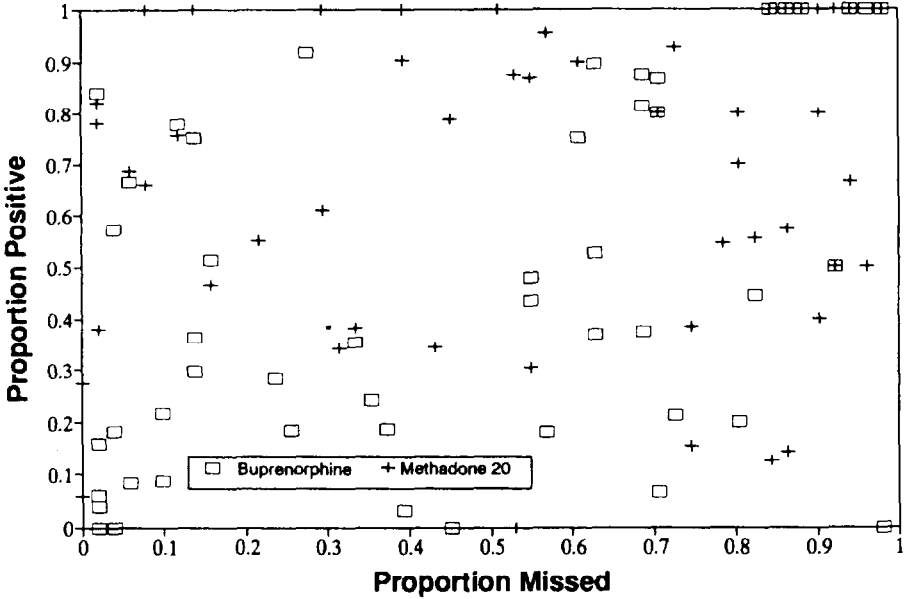
FIGURE 3. *Proportion of positive and missed tests, by subject*

proportion of positive tests. The difference in proportions of missingness, although not significant by itself, has a modest diluting effect.

The test based on the model for informative censoring provides the smallest p-value of all tests of the hypothesis $E[\theta_{i1}] = E[\theta_{i2}]$. It is substantially larger than the test based on the $\overline{Y}_{i \cdot k}$s. This is not surprising since tests that require more assumptions are generally more efficient.

The estimated models for the two groups are presented in table 3. Using the Wald statistics, the missing data are seen to be highly informative for the methadone 20 group and not as informative for the buprenorphine group. Since subjects with fewer missing observations tend to drop out later, it is somewhat misleading to talk about the separate effects of m_i and L_i. However, note that for the methadone 20 group, subjects with larger m_is (i.e., fewer missing observations) tend to have a lower proportion of positive tests. Subjects who drop out later are more likely to test positive.

108

The estimated average proportion of positive tests for the two groups in table 3 is quite close to the average of the c_is. However, the variance of $\sum_i \hat{E}[\theta_{ik}]/n_k$ is substantially smaller than $\hat{V}[\sum_i \overline{Y}_{i \cdot k}/n_k]$ from table 1. As mentioned previously, the smaller variability is not surprising since the model introduces more "structure" to the data. However, the ratio of the estimated variances is similar for the two approaches.

For simplicity, detailed comparison involved two groups. Finally, the authors use an evaluation of the three arms using the O'Brien rank test, which simultaneously tests equality of efficacy and missingness over all three arms. The Kruskal-Wallis chi-square test with two degrees of freedom had the value of 8.55 (p=.01). We then considered the three pairwise comparisons and used a Bonferroni correction to determine an (approximately normal) critical value of 2.39. Buprenorphine is better than methadone 20 (0_1 = -2.70), but not better than methadone 60 (0_1 = -.95). Methadone 60 was better than methadone 20, but not significantly so (0_1 = 2.15).

TABLE 1. *Some summary statistics for the buprenorphine and methadone 20 groups*

Statistic	Group	
	Buprenorphine (n=52)	Methadone 20 (n=55)
$\sum_i \overline{Y}_{i \cdot}/n$	.49	.69
$\hat{\tau}^2$	.11	.07
$\hat{V}[\sum_i \overline{Y}_{i \cdot}/n]$	.0025	.0016
$\sum_i \overline{D}_{i \cdot}/n$	.48	.58
Sample variance of $\overline{D}_{i \cdot}$	.1163	.1119

TABLE 2. *Tests comparing the response proportions between the buprenorphine and methadone 20 groups*

Test	Hypothesis	Z Value
Difference in average $\overline{Y}_{i \cdot k}$s with random effects variance	$E\,[\theta_{i1}] = E\,[\theta_{i2}]$	-3.05
Rank version of above with Empirical Bayes adjudication of ties		-3.01
Difference in average $\overline{D}_{i \cdot}$s		-1.63
Difference in average $\overline{Y}_{i \cdot k}$s with missing to 1 imputation		-3.48
O'Brien's rank test		-2.70
Difference in average $\hat{E}\,[\theta_{ik}]/n_k$s		-3.82

TABLE 3. *Parameter estimates for the models of informative censoring. The estimated mixing distribution for the buprenorphine (methadone 20) group had 4 (2) points of support. Estimated Wald statistics are provided in parentheses.*

	Group	
Effect	Buprenorphine	Methadone 20
$\hat{B}_k$	1.20	.38
$\hat{\beta}_{1k}$	-.022 (-1.29)	-.136 (-6.14)
$\hat{\beta}_{2k}$	-.021 (-1.14)	.131 (5.59)
$\sum_i \hat{E}[\theta_{ik}]/n_k$	.48	.68
Sample variance of $\hat{V}[\sum_i \hat{E}[\theta_{ik}]/n_k]$	.00133	.00116

DISCUSSION

This chapter briefly introduces and illustrates several techniques that may be useful for dichotomous repeated measures with a substantial proportion of missing data. A more rigorous evaluation would be useful before definitive recommendations are made. Nonetheless, several points can be offered.

The rank test with Empirical Bayes adjudication is an appealing procedure because it allows an unbiased robust comparison of two proportions as long as it is assumed that the probability of a positive test does not vary with j. The analogous test of means may be more substantially affected by $\overline{Y}_j$s based on few observations. Furthermore, the means test requires some structure to derive a variance estimate.

The simple imputation of missing to positive might be favored either if one felt that subjects were taking drugs on missed days or if a combined endpoint were thought reasonable.

The attraction of the model-based approach is that, in principle, it can provide a test of the original hypothesis. The disadvantage to this approach is in the implementation. Issues of covariate selection and model fit need to be explored. Additionally, optimization requires some care due to the possibility of local maxima and numerical stability. For example, in the methadone 20 group, it seemed that an additional support point at $-\infty$ would slightly increase the log likelihood. However, this point was not included due to numerical problems with the information matrix.

In general, there may be additional information to aid investigators in deciding how to deal with each specific missing datum. For example, some missing data might correspond to occasions when the subject was strongly suspected of using opiates. Such additional information can be incorporated into a combination procedure in which the imputation from missing to positive is made for a subset of the data. The other procedures discussed in this chapter could then be applied to the partially transformed data. However, it is important to recognize that no statistical procedure will improve the results from a trial with more than two-thirds of the endpoint data missing. Ultimately, the quality of evidence from such a trial is more like that of an observational study.

REFERENCES

Follmann, D.A., and Lambert, D. Generalizing logistic regression by nonparametric mixing. *J Am Stat Assoc* 84:295-300, 1989.

O'Brien, P.C. Procedures for comparing samples with multiple endpoints. *Biometrics* 40:1079-1087, 1984.

Pierce, D.A., and Sands, B.R. *Extra-Bernoulli Variation in Binary Data.* Technical Report 46. Corvallis, OR: Oregon State University, Department of Statistics, 1975.

Pocock, S.J.; Geller, N.L.; and Tsiatis, A.A. The analysis of multiple endpoints in clinical trials. *Biometrics* 43:487-498, 1987.

Sahlroot, J.T., and Pledger, G.W. "Monitoring Plasma Levels in Response to Monitored Plasma Concentrations: Can Unblinded Staff Adhere to Objective Criteria?" Unpublished manuscript, 1991.

Stiratelli, R.; Laird, N.; and Ware, J.H. Random effects models for serial observations with binary response. *Biometrics* 40:961-971, 1984.

Wu, M.C., and Bailey, K.R. Estimation and comparison of changes in the presence of informed censoring: Conditional linear models. *Biometrics* 45:939-955, 1989.

Wu, M.C., and Carroll, R.J. Estimation and comparison of changes in the presence of informative censoring by modeling the censoring process. *Biometrics* 44:175-188, 1988.

Wu, M.C.; Hunsberger, S.; and Zucker, D. Comparison of changes in the presence of censoring: Parametric and nonparametric methods. In: *Proceedings of the Biopharmaceutical Section of the American Statistical Association.* Alexandria, VA: American Statistical Association, 1991, pp. 291-299.

ACKNOWLEDGMENTS

AUTHORS

Dean Follmann, Ph.D.
Mathematical Statistician

Margaret Wu, Ph.D.
Mathematical Statistician

Nancy Geller, Ph.D
Chief

Biostatistics Research Branch
National Heart, Lung, and Blood Institute
National Institutes of Health
Federal Building, Room 2A-11
Bethesda, MD 20892

Summary of Discussion: "Issues in the Analysis of Clinical Trials for Opiate Dependence" by Follmann, Wu, and Geller

Ram B. Jain

Dr. Jack C. Lee of the National Institute of Child Health and Human Development, National Institutes of Health, who reviewed this paper, expressed concern about the missing-at-random assumption implicitly made by the authors in some of the models used by them. The same concern was expressed by many other participants at one time or another. The assumption of missing at random is questionable since a missed visit might be dependent on opiate abuse during the days just prior to missed visits. Dr. Follmann replied that the model-based imputation method presented by him did not require the assumption of missing at random. The parameters a_0 and a_1 will be zero if the data are missing at random.

I believe the distinction between a *missed* observation and a *censored* observation was lost during this discussion. An observation is considered to be *censored* when a subject *permanently* drops out of the study. For the censoring to be informative, *total* experience or abuse history of the subject till the time of censoring should play a role and should be investigated. A *missed* observation, on the other hand, is a *temporary* event. For a missed observation to be informative, a single or only a few events just prior to the missed visit should play a role.

Dr. Lee also made a number of other suggestions that can be incorporated in the models to describe the phenomenon of drug addiction. He suggested that cyclic effects introduced by, for example, the pattern of drug abuse and their relationship with missed visits can be incorporated in the models, and the covariates that may affect b_{ijk} should be included in the models. He suggested that the whole patient population may be divided into four or five relatively homogeneous strata, and these strata can then be separately analyzed. These analyses may not need the assumption of missing at

114

random. Dr. Lee was also of the view that the total study, for example, may be divided into three periods and that each of these periods may be studied separately. This might help in studying the issue of compliance. Finally, he thought a goodness-of-fit test was lacking from the presentation.

There was a rather involved discussion about imputation of missing values and the effect this may have on the inferences that are drawn. Dr. Hedayat was concerned about always imputing missing observations to a single value. Variations in patient characteristics across different areas may justify imputation to different values. Dr. Wright was concerned about one treatment being favored over another because of the specific imputation procedures used by the statistician. Dr. Fisher was in favor of some kind of sensitivity analysis where missing observations are imputed to different values in different treatment groups to describe various possibilities under a different set of imputed observations. However, Dr. Geller was of the view that you may come up with any conclusion when the missing (censored) data are as massive as in drug abuse trials.

Another suggestion was to consider some of the multiple imputation procedures used by Dr. Don Rubin, Harvard University. This might give a handle on the variability induced by the imputation procedure itself. It was pointed out that one of the natural sets to select for multiple imputation would be the entire history of the patient,

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Analysis of Clinical Trials for Treatment of Opiate Dependence: What Are the Possibilities?

Ram B. Jain

INTRODUCTION

One of the major primary outcome variables in clinical trials for treatment of opiate dependence is the frequency of drug abuse, that is, of opiates (primarily heroin), after therapy for opiate dependence has been initiated. Because the episodes of opiate abuse are not directly observable, an estimate for the frequency of opiate abuse is obtained from the urine samples collected with a prespecified frequency and tested for the presence of opiates and their metabolites. Hence, a data sequence of binary numbers for each addict is available for analysis. Analyses of these data present serious obstacles.

To obtain the "true" estimate of the frequency of opiate abuse, it will be necessary that each positive urine sample represent an independent episode of opiate abuse. However, depending on the amount of opiate consumed by an addict during a given episode, it will not always be true. Two or more consecutive positive urine samples *may* represent the same episode of opiate abuse. In other words, there is a probability that treatment effect will be confounded with the carryover from one positive urine to another. It is difficult to estimate carryover, because the probability of carryover for a given addict varies from day to day and among addicts from one addict to another because of differentials in drug-seeking behavior. Consequently, using the available information on the kinetics of opiates, the frequency of urine samples is selected in such a way that the probability of carryover is minimized and the probability of being able to detect an episode of opiate abuse is maximized; note that the probability of carryover is not entirely eliminated. This is the first obstacle in analyzing these trials.

In clinical trials among drug addicts, the dropout rate is unavoidably high, to the order of 80 percent in a placebo group. Also, even during the period the addicts stay in the trial, they miss about one in every five scheduled visits for

treatment. Hence, the number of missing or censored data points may be as much as or more than the number of available data points, which reduces the power of the statistical tests of hypotheses.

In 15- to 20-week trials, urine samples may be collected up to three times a week. Hence, each addict may have 50-or-so data points for analysis. As such, the problem of analyzing these trials may be perceived as a 50-or-so dimensional problem. The selection of a powerful statistical method that will permit 50-or-so dimensions with sample sizes on the order of 150 to 500 patients with substantial missing data to detect "true" treatment differences is a serious challenge.

Before the possibilities for analyzing these data are considered, it would be beneficial to understand the nature of treatment for opiate dependence. Agonist therapy for opiate dependence essentially constitutes *replacing* the abused opiate with another, most likely a synthetic opiate (called an opiate agonist or an opiate partial agonist), with relatively less potential for abuse. In *replacement* treatment, the next dose is given when the effect of the previous dose is about to *wear off.* If the next dose is not given *in time,* the addict is more likely to go out and seek the illegal drug of abuse. Overdosing amounts to exposing the addict to the addictive potential of the *replacement* opiate. Hence, each dose of the replacement opiate has its *own* pharmacological effect and may be considered as one *unit* of treatment.

According to Blaine and colleagues (1981), replacement therapy "is intended to . . . achieve a more pharmacologically stable physiological state."   Each *unit* of replacement therapy, if successful, should lead to a physiological state that is pharmacologically more stable than with the previous *unit* of replacement treatment. Hence, attainment of a *fully* pharmacologically stable physiological state at which the addict does not seek the abused opiate and is ready for *detoxification* is going to be a gradual, *one-step-at-a-time* process.

WHAT ARE THE POSSIBILITIES?

Let P_{ij} $(i = 0, 1, 2, . . ., m)$ be the probabilities (table 1) of an addict using the abused opiate before entering the trial $(i = 0)$ and after scheduled dose i $(i > 0)$ of the replacement opiate j. If each unit of the replacement opiate is consistently successful and has no "reverse" therapeutic effect, $P_{(i+1)j} \leq P_{ij}$, $i = 0, ... m - 1$. However, because of missed doses and other factors, P_{ij} can assume any value between 0 and 1. Also, a data point s after scheduled dose $s, s = 2, . . ., m,$ for a given addict may not be available because of the missed dose s or because he or she dropped out of the trial after dose $n, n < m.$

TABLE 1. *Probabilities of opiate abuse*

Dose	Urine Sample	Probability (Opiate Abuse) on Treatment j		
0	0	P_{0j}	P'_{0j}	P_{+0j}
1	1	P_{1j}	P'_{1j}	P_{+1j}
2		P_{2j}		
3	2	P_{3j}	P'_{2j}	P_{+2j}
.				
.				
m	n	P_{mj}	P'_{nj}	P_{+nj}

P_{ij} = probability of opiate abuse *after* administered dose i for treatment j
P'_{kj} = probability of opiate abuse *after* the last administered dose before
      scheduled urine sample i for treatment j
P_{+kj} = probability of urine sample k being positive for treatment j

Also, because urine samples are not collected after each dose, not all P_{ij} are estimable. For example, in the ARC 090 trial, urine samples were collected only after the Sunday, Tuesday, and Thursday doses, that is, on Mondays, Wednesdays, and Fridays. Thus, the number of doses of replacement opiate administered between consecutive urine samples varied from 1 to 2. Furthermore, because different addicts enter trials on different days of the week, the number of doses of replacement opiate administered between urine samples n and $n + 1$ varies from addict to addict.

Let P'_{kj} $(k = 0, 1, . . ., n)$ (table 1) be the probabilities of an addict using the abused opiate before entering the trial $(k = 0)$ and *after* the last administered dose before urine sample k $(k > 0)$ is scheduled to be collected for the treatment group receiving the replacement opiate j. Again, not all data points are available because of addicts missing one or more of the scheduled doses before the urine sample k is scheduled to be collected or because of not providing one or more scheduled urine samples, for example, for nonvisits or because of dropping out of the study after providing u $(u < k)$ urine samples.

The most practical way to estimate P'_{kj}. is to assay urine sample k for the presence of the abused opiate(s) and/or its (their) metabolites. However, because of many reasons, more fully described in Jain's chapter "Design of Clinical Trials for Treatment of Opiate Dependence: What Is Missing?" (this volume), the probability of a urine sample detecting an episode of opiate abuse depends on several factors, primarily the duration between the *last* episode of drug abuse and the time the current urine sample was obtained. Let P_{+kj}

(table 1) be the probability of a urine sample k for treatment j being declared as positive for opiate. Then, in an experiment, the best that can be done is to estimate P_{+kj} and hope that P_{+kj} is the best available estimate of P'_{kj}.

There are at least three distinct possibilities for analyzing these data or estimating P_{+kj}. First, reduce the multiple data points for each addict to one and then use regular inference procedures to compare the efficacy of different treatments. For example, multiple data points obtained from urine samples for an addict may be reduced to a single data point defined as the proportion of positive urines, or alternatively, his or her overall profile/pattern of +/- urines can be classified by some rank order procedure as a single rank. Let this possibility be denoted as DATA-REDUC-1.

If sequential performance of successive *units* of replacement opiate is of interest, estimates of P_{+kj}'s can be obtained, trends studied, and a summary statistic $\Sigma w_{kj} P_{+kj}$ obtained to evaluate the program performance of the different treatments. Weights w_{kj} can be defined in many different ways and are well documented in statistical literature. Let this possibility be denoted as ANAL-SEQ-UNIT.

It will be in order here to clarify the major distinction between the summary statistic obtained from ANAL-SEQ-UNIT and the single statistic obtained from DATA-REDUC-1 procedures. Whereas the summary statistic obtained from ANAL-SEQ-UNIT procedures is adjusted for differentials in treatment performances and sample sizes over time, the single statistic obtained from DATA-REDUC-1 basically ignores these differences in treatment performances and sample sizes over time. However, the latter is simpler to compute and understand.

Also, attention can be focused on only the positive urines, and a correlational structure between time to various positive urines or failures can be studied. In other words, data can be analyzed as a multiple failure problem. Let this possibility be denoted as MULT-FAIL.

Some of these possibilities were explored in analyzing the ARC 090 data for buprenorphine vs. methadone 60 mg treatment, and some of these results are presented.

DATA-REDUC-1

If the multiple data points are reduced to one data point for each addict, one of the first temptations would be to use some form of parametric or nonparametric analysis of variance. However, censored observations are not permitted in

119

analysis of variance, and then what is to be done with missing observations? Both missing and censored observations can be considered as "negative" or "positive," as can some other combination of "negative" and "positive," probably depending on the reason for missing and censored observations. But then there are at least as much "made-up" data as the real observed data. This is probably not acceptable to most analysts.

If the proportion of positive urines, p_{+kj}, is to be computed for addict $v, v = 1, . . ., n_j$ in treatment jfor using parametric analysis of variance, the censored and missing observations may be excluded from the analysis; that is, a different denominator is used for each addict. This would violate the assumption that each subject in a given treatment group is drawn from the same population. In addition, since the probability p_{+kj} of a positive urine for urine sample k varies with $k,$ the *single* variable y denved from multiple data points will be the sum of u binomial variables with parameters $n = 1$ and $P = P_{+tk}$. Is y normally distributed and with what parameters? However, irrespective of theoretical objection, this possibility was explored for the ARC 090 data, and the results are given in table 2. No significant differences were observed.

An additional problem with both parametric and nonparametric analysis of variance is that information about the pattern of positive and negative urines or temporal correlations is lost. Also, since the kinetics of the drugs are different, the information about the relationship between drug effect and time is lost.

Survival methods that permit censored observations can be used with a little more confidence. However, in addition to the problem of missing observations, the definition of what constitutes a failure may be subjective, and depending on the definition used, the power of statistical procedure may become too low

TABLE 2. *Parametric analysis of variance results for the ARC 090 study (maintenance period only)*

Treatment Group	Missing Values Treated as	N	Mean p_{+kj} (SD)	t	p
Buprenorphine	Negative	48	.33(.26)	-.46	.65
	Positive	48	.54(.33)	-.85	.40
	Excluded	47	.45(.35)	-.29	.78
Methadone 60 mg	Negative	51	.35(.30)		
	Positive	51	.60(.33)		
	Excluded	47	.48(.36)		

(because of too few failures) or the trial may be over too soon, thus making most of the data observed unused. For example, if the first positive urine is used as a measure of treatment failure, ARC 090 would probably be over in a week or so. For ARC 090, two consecutive Monday positive urines were used, starting with the fourth Monday of treatment as the measure of treatment failure; the results are given in table 3. No significant differences were found. Kaplan-Meier survival curves are displayed in figure 1. The number of *failures* in each group was 25.

Another measure of treatment *failure,* that is, the beginning of first drug-free period of 28 days or more, was also used. The number of failures using this criterion was 13 in the buprenorphine group and 7 in the methadone 60 mg group. The results are given in table 4, and Kaplan-Meier curves are plotted in figure 2. As can be seen from table 4, the two statistics can give different results. Only the Breslow statistics provide significant results. Hence, depending on the definition of a failure, different methods of inference can give different results.

Another possibility is being explored by Dr. John Harter, director of the Pilot Drug Evaluation Division of the Food and Drug Administration, in analyzing analgesic trials that use a combination of a *sorting routine* and a nonparametric rank sum test. The sorting routine first sorts *all* subjects by their pain intensities at time (sample) 1; then each distinct subgroup obtained after first sort is sorted by its pain intensity at time (sample) 2, and so on. After the last sort, the subjects are ranked according to their profiles in ascending or descending order. Then, ranks are summed for each treatment group, and a rank sum test to evaluate treatment differences may be used. This approach may also be tried for drug abuse trials.

TABLE 3. *Results of survival analysis of the ARC 090 study (maintenance period only) using two consecutive Monday positive urines starting with the fourth Monday of treatment as treatment failure*

95-Percent Brookmeyer-Crowley Confidence Intervals in Days for Median Survival Time for		Mantel-Cox Chi-Sq *(p)*	Likelihood Ratio Chi-Sq *(p)*
Buprenorphine	Methadone 60 mg		
48.0-90.0	35.0-77.0	0.50 (.48)	.09 (.77)

```
       1.0   +H                                                            +
             -H.............                                               -
             -           H.                                                -
             -           B.                                                -
             -           ..                                                -
             -           H.                                                -
             -            .                                                -
             -            .                                                -
        .80  +           ..                                                +
             -           .B..                                              -
             -           H  B.                                             -
             -           .  B                                              -
             -           .  B...                                           -
             -           H.                                                -
             -            .    B.                                          -
             -            H                                                -
  .60  +            .      B                                      +
             -           H..   B...                                        -
             -            H.    B                                          -
             -            H.......                                         -
             -               H.                                            -
             -               H.......                                      -
             -                 .   B...                                    -
        -                     H..   .                                   -
        .40  +                 H..  .                                      +
             -                 H...........                                -
             -                  H..  B........                             -
             -                       .  .     .                            -
             -                     H.................                      -
             -                            B......                          -
             -                                  H,                         -
        .20  +                                                            +
             -                                                            -
             -                                                            -
             -                                                            -
             -                                                            -
             -                                                            -
             -                                                            -
        0.0  +                                                            +
             .+....+....+....+....+....+....+....+....+....+....+....+.
               10.    30.     50.     70.     90.    110.
              0.0    20.     40.     60.     80.    100.    120.
```

(vertical label at left) Proportion Surviving

Time (in days) to Two Consecutive Monday Positive Urines After Fourth Monday
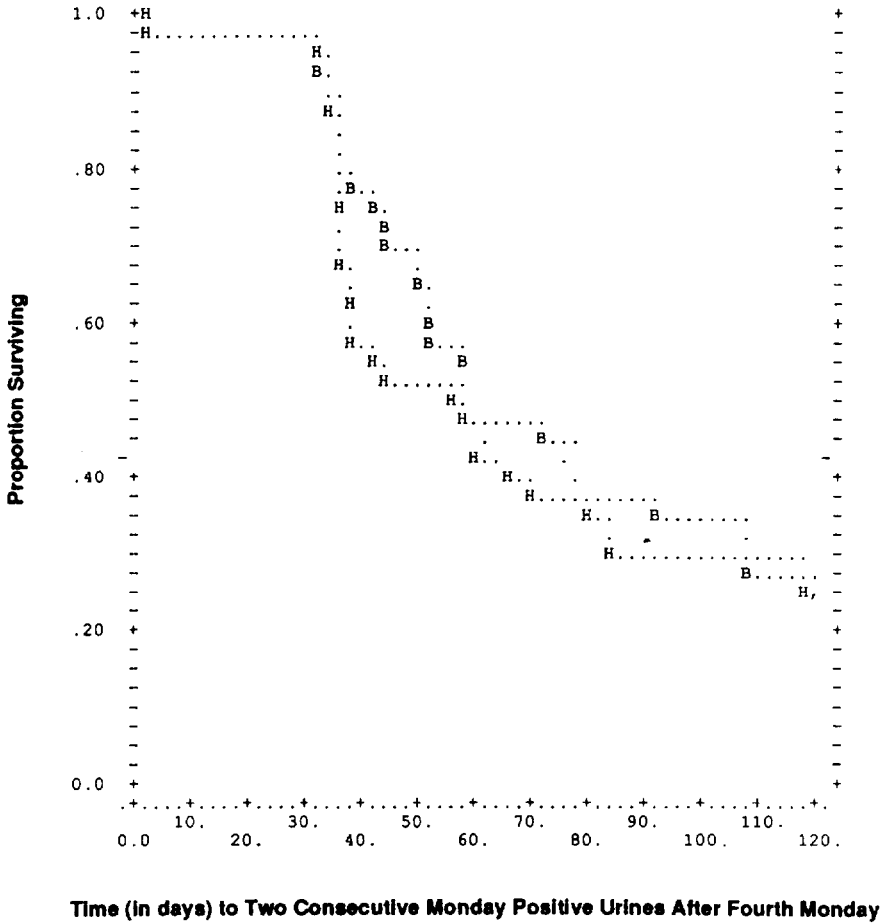
FIGURE 1. *Kaplan-Meier survival curves for ARC 090 study*

B = buprenorphine, H = methadone 60 mg

Consider six addicts on two different treatments who have their urine results on three samples as shown in table 5a. The subjects are first sorted according to their results on sample 1 as shown in table 5b. Subjects 1, 3, 5, and 6 have positive urines and as such are subgrouped first, followed by subjects 2 and 4, each of whom has a negative urine. If there were to be more than two distinct scores, the procedure would be the same. After the first sort, the

TABLE 4. *Results of survival analysis of the ARC 090 study (maintenance period only) using the beginning of the first drug-free period of 28 days or more as a treatment failure*

Mantel-Cox Chi-Sq *(p)*	Breslow Chi-Sq *(p)*	Likelihood Ratio Chi-Sq *(p)*
2.95 (.09)	5.51 (02)	2.94 (.09)
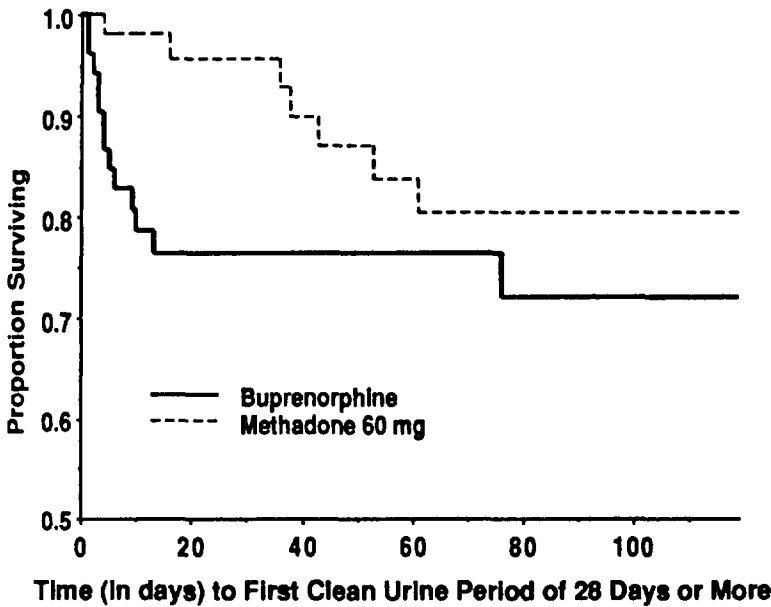
FIGURE 2. *Kaplan-Meier curves for ARC 090 study when treatment "failure" is defined as the first drug-free period of 28 days or more*

distinct subgroup of subjects 1, 3, 5, and 6 is sorted first by their results on urine sample 2, and then the second distinct subgroup of subjects 2 and 4 is sorted by their results on urine sample 2. This creates three distinct subgroups: subjects 3 and 6 with a (+,+) profile; subjects 1 and 5 with a (+,-) profile; and subjects 2 and 4 with a (-,+) profile (see table 5c). As shown in table 5d, each of these three subgroups is then sorted by results on the third urine sample, thus creating six distinct subgroups of subjects: subject 6 with a profile (+,+,+) ranked 1, subject 3 with a profile (+,+,-) ranked 2, subject 5 with a profile (+,-,+) ranked 3, subject 1 with a profile (+,-,-) ranked 4, subject 2 with a profile (-,+,+)

TABLE 5a. *Results of three urine samples from a hypothetical trial*

		Results of Urine Sample		
Patient Identification	Treatment Received	1	2	3
1	A	+	-	-
2	A	-	+	+
3	A	+	+	-
4	B	-	+	-
5	B	+	-	+
6	B	+	+	+

TABLE 5b. *Results from a hypothetical trial after first sort*

Patient Identification	Treatment Received	Results After First Sort
1	A	+
3	A	+
5	B	+
6	B	+
2	A	
4	B	

TABLE 5c. *Results from a hypothetical trial after first two sorts*

Patient Identification	Treatment Received	Results After First Two Sorts	
3	A	+	+
6	B	+	+
1	A	+	-
5	A	+	-
2	A	-	+
4	B	-	+

TABLE 5d. *Results from a hypothetical trial after three sorts*

Patient Identification	Treatment Received	Results After Three Sorts			Rank
6	B	+	+	+	1
3	A	+	+	-	2
5	B	+	-	+	3
1	A	+	-	-	4
2	A	-	+	+	5
4	B	-	+	-	6

ranked 5, and subject 4 with a profile (-,+,-) ranked 6. For subjects with the same profiles, average ranks can be calculated. The sum of ranks for treatment A = 11 may then be compared with sum of ranks for treatment B = 10.

There are two problems with this approach. First, as with any nonparametric rank test, the magnitude of treatment differences on original measurement scale is not available. Second, this approach puts too much weight on the first observation. For example, subject 3 with a profile of (+,+,-) is given the rank of 2, whereas subject 2 with a profile of (-,+,+) is given the rank of 5. Both the subjects have two out of three positives, but they are considered (almost) opposite extremes in this approach. However, it may be possible to come up with certain variations of this approach that do not rely on the first observations so heavily. For example, if clinically acceptable, the first few results may be ignored or a ranking mechanism may be developed based on some combination of result profiles and number of positives.

ANAL-SEQ-UNIT

The first approach to explore this possibility would be to construct 2x2 tables for each urine testing opportunity and compute, for example, a Mantel-Haenszel z-statistic (Mantel and Haenszel 1959; Miller 1981). For the ARC 090 study, the z-scores are displayed in figures 3, 4, and 5 when missing values are considered as missing, negative, and positive, respectively. A consistent pattern of superiority of buprenorphine over methadone 60 mg is observed, except probably during the middle of the study period. However, the degree of superiority of buprenorphine seems to be decreasing over the first half of the study period and then increasing again during the second half of the study period. Some, but not all, of this is explainable based on the process of self-selection as the study progresses and because of different sample sizes at different times during the study period. The differences probably lie in different kinetics of buprenorphine and methadone. Methadone is probably catching
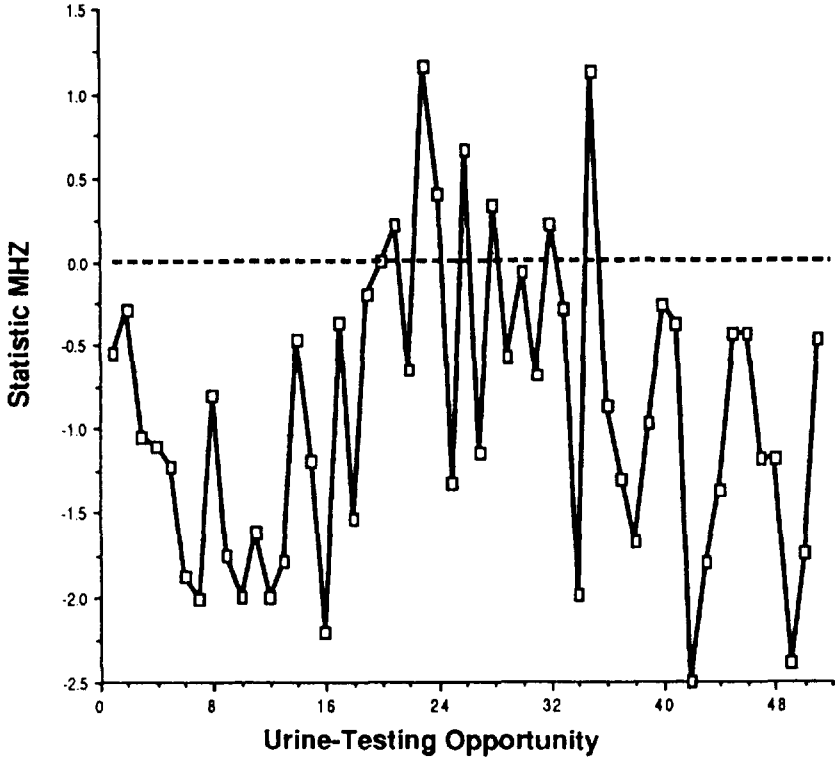
FIGURE 3. *Mantel-Haenszel z-statistic for each urine testing opportunity when missing values are considered as missing*

up with buprenorphine during the first half of the study period, as may be suggested from percent of positive urines at different times during the study, as seen from figure 6.

However, a summary Mantel-Haenszel statistic cannot be validly calculated from individual 2x2 tables because this summary statistic does not account for correlations between individual 2x2 tables. How then can one calculate a summary index of program effectiveness? First, a simple though somewhat questionable alternative may be to score the direction of the relative efficacy of the two drugs for each time point or urine testing opportunity and use a binomial test to evaluate if, "overall," one drug is more effective than the other. Another alternative may be to use a weighted summary statistic for correlated tables as described in Wei and Johnson (1985). However, the use of, for example, a
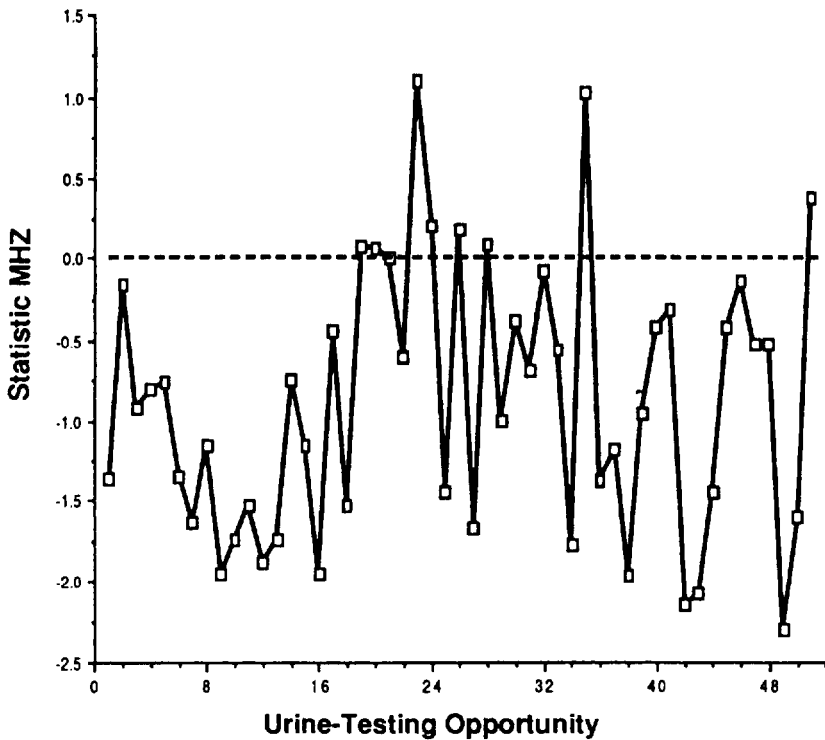
FIGURE 4. *Mantel-Haenszel z-statistic for each urine testing opportunity when missing values are considered as negative*

51-dimensional variance-covariance matrix as in the ARC 090 study with the data as sparse as they are, particularly during the last few weeks of the study, would certainly lead to some problems. For example, to solve such a huge and sparse matrix will be numerically difficult, and the dimension of the problem will adversely affect the power of the statistic.

The use of parametric repeated measure of analysis is even more problematic. In addition to the inadmissibility of missing and censored observations, the degree of robustness of repeated measure analysis of variance to analyze binary data is unknown when there are as many repeated *measures* as in these studies. Also, these studies do not generate traditional repeated measures data. Each of the two consecutive repeated measures is interrupted by the administration of the replacement opiate and possibly the use of opiate of abuse. This is different when compared with using a new instructional

127
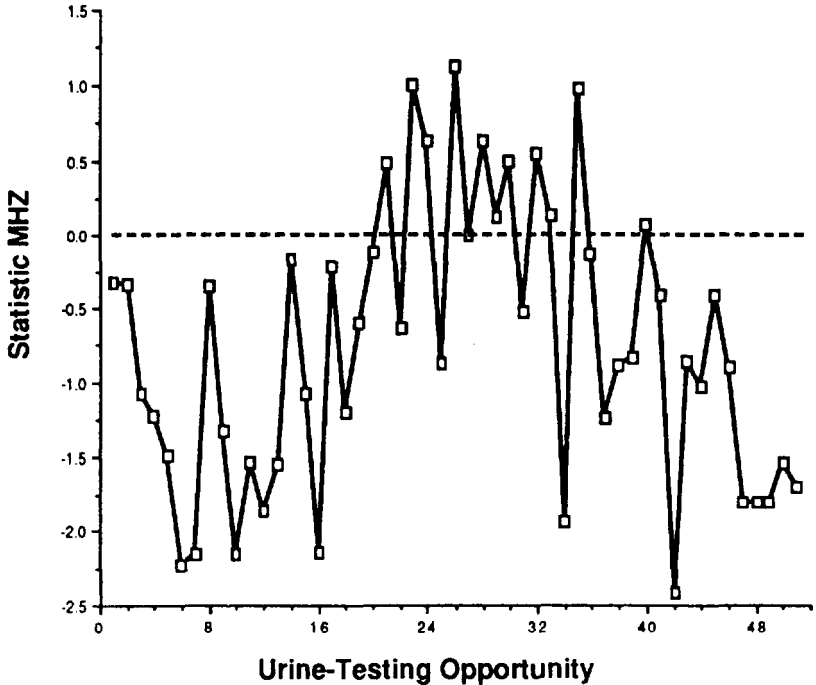
FIGURE 5. *Mantel-Haenszel z-statistic for each urine testing opportunity when missing values are considered as positive*

method for several months and comparing the effects of the traditional and the new method over a period of time. It is not certain if the repeated measure theory is applicable to these data. At best, these data seem to be multiply interrupted time series data.

MULT-FAIL

Several authors have considered the problem of analyzing multiple failures under various configurations (e.g., failures of the same type over time or failures of different types at a fixed point in time and space). Recent work in this area has been done by several researchers (Lagakos et al. 1978; Hsieh et al. 1983; Prentice et al. 1981; Gail et al. 1980; Lawless 1987; Wei and Lachin 1984; Thall and Lachin 1988; Wei and Stram 1988; Wei et al. 1989; Lin 1990). Some of these authors, for example, Wei and Stram (1988) and Wei and colleagues (1989). used a regression-based approach, whereas others, such as Wei and
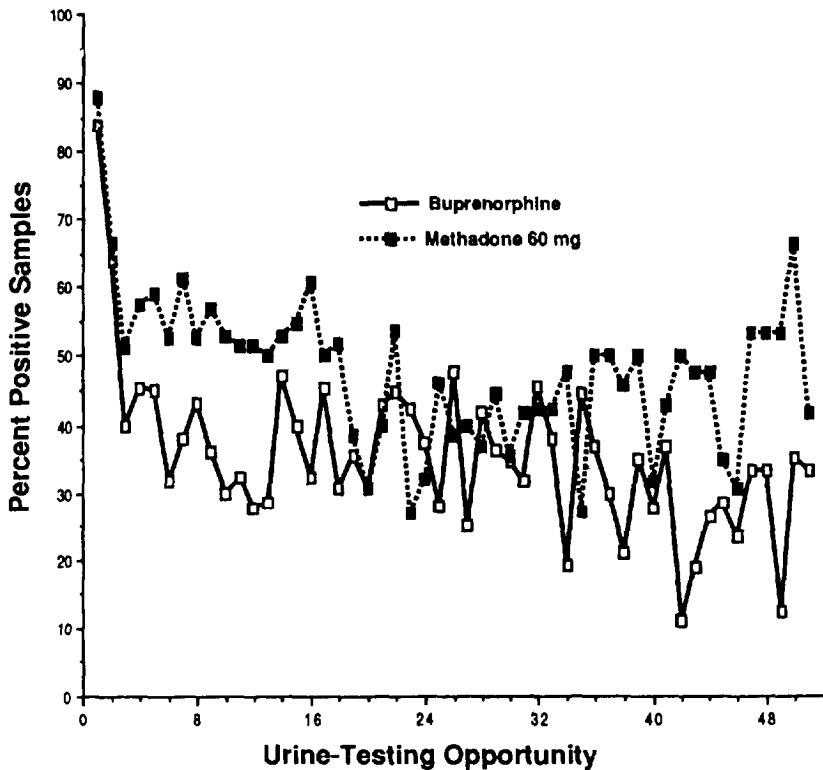
FIGURE 6. *Percentpositive urines for ARC 090 study*

Lachin (1984) and Thall and Lachin (1988), used multivariate versions of log rank (Mantel 1966) and/or the Gehan test (Gehan 1965) to analyze multiple failures. The regression approach of Wei and colleagues (1989), which is a multivariate version of Cox's proportional hazards model (Cox 1972), imposes the least restrictive structure on recurring events (failures) and thus is very appealing.

The regression approach of Wei and colleagues (1989) will be an excellent choice if the number of failures in the model is limited and there are many subjects in the study. In the ARC 090 study, the number of subjects, 162 across three treatment groups, was probably sufficient to use this model, but each subject could also experience up to 51 failures in the 17-week maintenance phase of the study. Hence, to use this model, there was no choice but to use an algorithm that reduces the maximum number of failures

to 17. Even though this approach does permit censored observations, the missing observations must still be handled in some way. In fact, the algorithm used to reduce 51-dimensional data to 17-dimensional data more or less solved this problem, except when all three samples during a week were missing.

A weekly index was developed for urine samples being positive or negative for opiates. If at least one of the three samples was positive or all samples for a given week were missing, that week was considered to be positive for opiates. Otherwise, that week was considered to be negative for opiates. Thus, the maximum number of failures was limited to 17 for this analysis. However, to avoid too many ties, the time (in days) to each failure used to compute various statistics was defined as the time to first positive urine or missing observation (if all observations were missing during a week) during the week in consideration.

This algorithm does result in some loss of information, for example, one who has three positive urines during a week is treated the same way as one who has only one positive urine during that week. Hopefully, this loss of information will be *random* and *uniform* across different treatment groups and will result in a valid comparison. No formal statistical tests were done to verify this.

Only one covariate-that is, treatment assignment-was used for analyzing ARC 090 data (1 = buprenorphine, 0 = methadone 60 mg). Thus, 17 regression coefficients, one for each week, were estimable. A joint test of hypothesis testing $H_k:\beta_k = 0$, k = 1, . . ., 17 was conducted. An estimate of common regression coefficient, $\Omega = \Sigma c_i\beta_i$, i = 1, . . .,17 was also obtained and tested for $\Omega = 0$. The weights c_j were optimally calculated by the program MULCOX (Lin 1990). A negative regression coefficient indicates a decreased hazard rate for buprenorphine compared with methadone 60 mg, that is, a negative regression coefficient favors buprenorphine treatment. Also, a hazard ratio of less than one favors buprenorphine treatment. The hypothesis $H_k:\beta_k = 0$ was not rejected (Wald statistic with 17 degrees of freedom = 20.89, *p* = .23). However, the estimate of $\Omega = \Sigma c_j\beta_j$ was found to be significantly different than zero (Ω = -0.294, p = .04), indicating an "average" superiority of buprenorphine over methadone 60 mg. The 95-percent confidence interval for the common hazard ratio of .746 was (.566, .983).

The hazard ratios for each week are plotted in figure 7 indicatlng consistent superiority of buprenorphine.
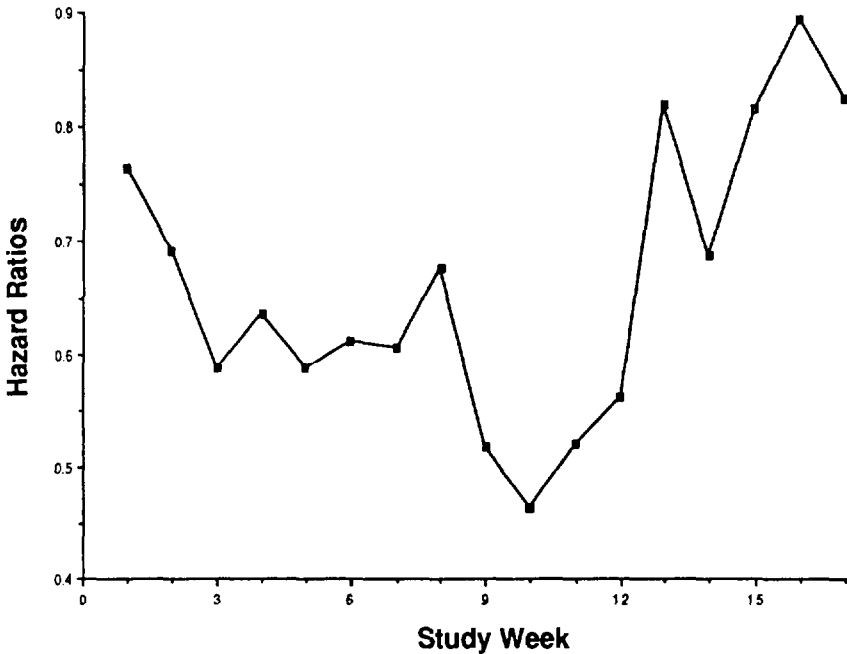
FIGURE 7. *Hazard ratios for each week for ARC 090 study*

WHAT ARE THE PROBLEMS?

The biggest problems in analyzing these data are:

1. The order of dimension (51-dimensional)
2. The sparseness of data
3. The problem of missing values

None of these problems seems to be handled too well by any of the possibilities
explored in this chapter. DATA-REDUC-1 methods do reduce the data to
one dimension but at a tremendous cost-complete loss of information about
correlational structures between various dimensions and more or less no
ability to handle missing and/or censored observations. In fact, some of the
DATA-REDUC-1 methods make no distinction between missing and censored
observations. ANAL-SEQ-UNIT methods do handle one dimension at a time
but have difficulty combining information unless some sort of *miniature* data
reduction scheme can be implemented. MULT-FAIL methods do handle
censored data, but heavy censoring causes loss of power, and dimension

of data must be reduced somewhat by using a *miniature* data reduction scheme. However, the possibly informative nature of censoring causes interpretational difficulties. The sparseness is either ignored or subjectively handled in DATA-REDUC-1 and ANAL-SEQ-UNIT methods.

None of the methods has any ability to handle missing values without outside intervention. The solution may be to consider a missing observation as the third stage as discussed by Weng (this volume). Another possibility suggested by Dr. Gross of the Medical University of South Carolina is to consider a quadrinomial model with four categories—positive, negative, missing, and censored-and then to consider a conditional binomial model in which conditioning is on the later two categories.

OTHER PRIMARY VARIABLES AND THEIR ANALYSES

One of the other three primary variables of interest in these clinical trials is the retention rates in the treatment program. These data can easily be analyzed by any one of the survival analytic techniques. However, the problem of informative dropouts may have to be handled in some way. Some of the work in this area is due to Dr. Margaret Wu of the National Heart, Lung, and Blood Institute (Wu and Bailey 1988, 1989; Wu and Carroll 1988).

One of the self-reported measures of drug abuse is the "craving" scores obtained periodically during the course of the study. Before entry into the trial (time 0) and at times $i, i = 1, \ldots, m,$ addicts are asked to report how much craving or need or desire they had during the last few days (e.g., a week or since the last time they visited the clinic) for the abused drug. Usually they are asked to "mark" the intensity of their craving or need on a 100-mm-long line called a craving scale, such as the one shown in figure 8. A score of zero means no craving, and a score of 100 means the most intense craving *ever* experienced. Let S_{ij}, be the craving scores reported by an addict on treatment j at time $i, i = 0, 1, \ldots, m.$ These data, like the urine data, have missing and censored observations. There are several ways to analyze these data. Regression analysis can be performed on either $S_{ij} (i = 0, \ldots, m)$ or on $S_{ii} - S_{0j} (i = 1, \ldots, m),$ and standard tests for ß = 0 or $ß_j-ß_k, = 0$ can be performed. Alternatively, regression analysis for multiple failures based on proportional hazards model such as those described by Wei and colleagues (1989) can also be used.

Another important outcome variable in the drug abuse trials IS the physician's (or staff's or patient's) global impression of an addict's status with respect to his or her drug-seeking behavior at different time points during the study as compared with a previous time point or compared with his or her status at the

FIGURE 8. *Craving scale used in drug abuse research*

time of entry into the study. Generally, these physician's (or staff's or patient's) scores are obtained on a 3- to 5-point rating scale. These data can be analyzed the same way as craving scores data, or in addition, change in the status can be evaluated by a one-sample or two-sample (Feuer and Kessler 1989) McNemar's chi-square test statistic.

REFERENCES

Blaine, J.D.; Thomas, D.B.; Barnett, G.; Whysner, J.A.; and Renault, P.F. Levo-alpha acetylmethadol (LAAM): Clinical utility and pharmaceutical development. In: Lowinson, J.H., and Ruiz, P., eds. *Substance Abuse Clinical Problems and Perspectives.* Baltimore/London: Williams & Wilkins, 1981. pp. 360-388.

Cox, D.R. Regression models and life tables (with discussion). *J Royal Stat Soc B* 34:187-220, 1972.

Feuer, E.J., and Kessler, L.G. Test statistic and sample size for a two-sample McNemar test. *Biometrics* 45:629-636, 1989.

Gail, M.H.; Santner, T.J.; and Brown, C.C. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 36:255-266, 1980.

Gehan, E.A. A generalized two-sample Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52:203-223, 1965.

Hsieh, F.Y.; Crowley, J.; and Tormey, D.C. Some test statistics for use in multistate survival analysis. *Biometrika* 70:111-119, 1983.

Lagakos, S.W.; Sommer, C.J.; and Zelen, M. Semi-Markov models for partially censored data. *Biometrika* 65:311-317, 1978.

Lawless, J.F. Regression methods for Poisson process data. *J Am Stat Assoc* 82:808-815, 1987.

Lin, D.Y. MULCOX: A computer program for the Cox regression analysis of multiple failure time variables. *Comput Methods Programs Biomed* 32:125-135, 1990.

Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50:163-170, 1966.

Mantel, N., and Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 22:719-748, 1959.

Miller, R.G., Jr. *Survival Analysis.* New York: Wiley, 1981.

Prentice, R.L.; Williams, B.J.; and Peterson, A.V. On the regression analysis of multivariate failure time data. *Biometrika* 68:373-379, 1981.

Thall, P.F., and Lachin, J.M. Analysis of recurrent events: Nonparametric methods for random-interval count data. *J Am Stat Assoc* 83:339-347, 1988.

Wei, L.J., and Johnson, W.E. Combining dependent tests with Incomplete repeated measurements. *Biometrika* 72:359-364, 1985.

Wei, L.J., and Lachin, J.M. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *J Am Stat Assoc* 79:653-661, 1984.

Wei, L.J.; Lin, D.Y.; and Weissfeld, L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 84:1065-1073, 1989.

Wei, L.J., and Stram, D.O. Analyzing repeated measurements with possibly missing observations by modelling marginal distributions. *Stat Med* 7:139-148, 1988.

Wu, M.C., and Bailey, K. Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Stat Med* 71:337-346, 1988.

Wu, M.C., and Bailey, K.R. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* 45:939-955, 1989.

Wu, M.C., and Carroll, R.J. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 44:175-188, 1988.

ACKNOWLEDGMENT

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Summary of Discussion: "Analysis of Clinical Trials for Treatment of Opiate Dependence: What Are the Possibilities?"

Ram B. Jain

During my talk I made a comment about the difficulty of explaining certain statistical methods to clinicians. Dr. Gorodetzky thought the likelihood of explaining the details of some of the statistical analysis to clinicians is very remote. If the statisticians can agree on what is an appropriate method of analysis, a qualitative discussion or description of the method along with discussion of results in relation to analysis would be sufficient. For Dr. Fisher, war was too important to be left to generals. He would not hesitate to speak on clinical matters, and sometimes the best statistical ideas do come from the clinicians. Dr. Gorodetzky remarked,

> I think the real problem is sometimes we tend not to talk each other's languages, and we tend to be way out here clinically and way out here statistically. If we can come a little bit more towards the middle with a little bit of mathematical understanding from a clinician and a little bit of clinical understanding from the statistician, there can be a very productive interchange.

Dr. Geller found pictures (e.g., cumulative hazard plots) to be very useful in helping clinicians understand some complicated statistical concepts. One should not try to explain every little detail because it is really not important to clinicians.

Dr. Geller found ranking methods to be a rich tool for analyzing multiple endpoints data (e.g., proportion positive and proportion missing) also. However, there may be some price to pay (e.g., loss in power) when parametric methods are applicable but nonparametric procedures are used. In addition, magnitude of treatment effects is not easily discernible when ranking methods are used. There are ways to go back to the original

unranked data, but they do not always work and may not always be desirable. Dr. Fisher did not think one should necessarily be tied to description of, for example, magnitude of treatment effect going along precisely with the specific test of hypothesis used to compute p values.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Toward a Dynamic Analysis of Disease-State Transition Monitored by Serial Clinical Laboratory Tests*

T.S. Weng

INTRODUCTION

In many clinical trials dealing with the monitoring and/or management of a chronic disease under medical treatment, it is customary to follow each patient up to some censoring time. The observations usually consist of longitudinal counts of patients in cohorts with common disease states identified by an ad hoc laboratory test repeatedly administered over a fixed sequence of time points. For example (see Jain's chapter, "Analysis of Clinical Trials for Treatment of Opiate Dependence: What Are the Possibilities?," this volume), in a randomized clinical trial (ARC 090) to evaluate the efficacy of buprenorphine for the treatment of opiate addiction, 162 qualified patients were put through a 17-week maintenance phase in three separate treatment groups: Group 1 was maintained on 8 mg of buprenorphine administered sublingually daily, and groups 2 and 3 were maintained on 20 mg and 60 mg, respectively, of methadone (positive control) administered orally daily. To evaluate the frequency of opiate abuse, all patients were asked to provide urine samples three times weekly on Mondays, Wednesdays, and Fridays. These samples were assayed to detect the presence of opiates (mainly heroin or morphine). A positive sample was defined as a possible treatment failure. Due to missed clinic visits or other reasons, 19.8, 17.7, and 17.7 percent, respectively, of urine samples from the three treatment groups were uncollected. Furthermore, the percentages of patients lost to followup in these groups were noted to run up to 60.4, 80.0, and 63.0 percent, respectively. If it were not for the massive, possibly nonrandom missing observations and loss of patients to followup, as well as for the seemingly time-dependent nature of the data encountered, this clinical trial could have been analyzed by the popular method of survival analysis using either the multivariate versions of Gehan's log-rank tests (Gehan 1965; Peto and Peto 1972; Wei and Lachin 1984) or the generalized versions

of Cox's semiparametric, proportional hazard model for censored failure time with covariates acting as treatment responses (Prentice et al. 1981; Gail 1981; Wei et al. 1989; Lin 1990).

The purpose of this chapter is to propose a stochastic compartmental model as an alternative for modeling the data generated from the aforementioned study, thereby evaluating the efficacy of buprenorphine against methadone in the treatment of opiate addiction. The plan of this chapter is as follows: First, a closed three-compartment system is introduced by which patients are classified according to their patterns and directions of response to medication during the course of treatment. This is followed by the introduction of a Markov process, which provides a natural context for addressing the problem of statistical dependence among successive observations and characterizes the dynamics of disease-state transition within the compartmental system. Based on the assumption that this synthesized stochastic compartmental model is piecewise stationary in time, an iterative weighted conditional nonlinear least-squares procedure is then developed to facilitate parameter estimation. The results are then applied to analyze the ARC 090 study to draw conclusions on the efficacy of buprenorphine treatment. Finally, a general discussion is given.

COMPARTMENTAL MODEL

The patient pool in the ARC 090 study can be partitioned into three cohorts or compartments (see figure 1 below) that are each numbered 1, 2, or 3 depending on whether they encompass patients who have tested negative (-) for opiates, positive (+) for opiates, or have missed the test with the potential for being lost to followup (M/L). In figure 1, the compartments are represented as boxes with arrows between boxes indicating the direction of disease-state transitions, Let N be the total number of patients and $N_i(t)$ be the number of patients in compartment i (i = 1,2,3) at time t > 0 and let $\lambda_{ji}(t)$ denote the transition rate from compartment i to compartment j (i,j = 1,2,3) at time t > 0. Patients will then be included in different compartments depending on the results of their urinary tests or on whether they comply with the urinary test schedule. This compartmental system, within which all compartments (or states) communicate with one another, is regarded as closed in the sense that $\sum_j N_j(t) = N$ at any time t > 0. The individual patients in the system are assumed to act independently without being influenced by others.

STOCHASTIC PROCESS

The dynamics of changes in disease states within this system may be described by a Markov process {X(t): 0 < t ∞} defined on the state space **S** = {1, negative; 2, positive; 3, missing or lost to followup} with the associated transition
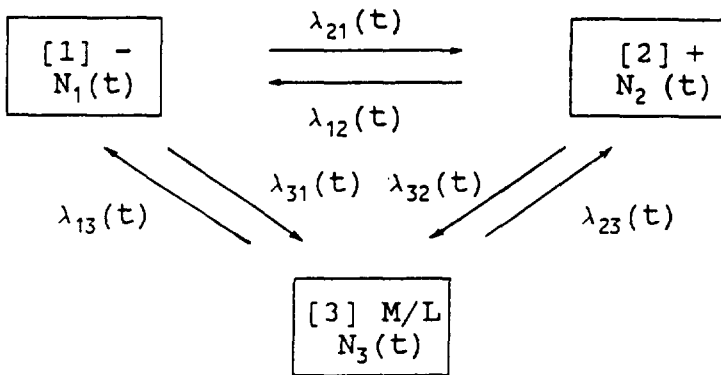
FIGURE 1. *Schematic diagram of a closed three-compartment model with $\lambda_{ji}(t)$ representing rates of transition between pairs of compartments at time $t > 0$ (i,j = 1,2,3)*

probability matrix (or transition matrix, for brevity) $\mathbf{P}(t,t_0) = [P_{ji}(t,t_0)]$, $0 \le t_0 \le t$ and the transition rate matrix (or rate matrix, for brevity) $K(t) = [\lambda_{ji}(t)]$, i,j ϵ **S**, where

$$P_{ji}(t, \ t_0 = \text{Pr } \{X(t) = j | X(t_0) = i\}, \tag{3.1}$$

$$0 \le \lambda_{ji}(t) = \lim_{\delta t \downarrow 0} P_{ji}(t + \delta t, t)/\delta t, \ j \ne i, \text{ and } \lambda_{ii}(t) = -\Sigma_{j \ne i}\lambda_{ji}(t). \tag{3.2}$$

For the time being, let $\lambda_{ji}(t) = \lambda_{ji}$ for all i,j so that $\{X(t)\}$ becomes a stationary process with $\mathbf{P}(t,t,)$ being a function of $t - t_0$, only. Without loss of generality, therefore, it may be assumed that $t_0 = 0$, so one can simply write $\mathbf{P}(t) = \mathbf{P}(t,0)$ Under these assumptions, the transition matrix P(t) is uniquely given by the Kolmogorov forward differential equation

$$(d/dt)\mathbf{P}(t) = \mathbf{K}\mathbf{P}(t) \tag{3.3}$$

with the initial condition $\mathbf{P}(0) = \mathbf{I}$, a 3x3 identity matrix. In the above equation, the rate matrix **K** is singular, as can be seen by the expressions in (3.2). Thus, the eigenvalues of **K** are given by $0, -\alpha$, and $-\beta$ ($0 < a, \beta < 1; \alpha \ne \beta$ where

$$\alpha, \beta = (1/2)(\lambda_1 + \lambda_2 + \lambda_3) \pm (1/2)\sqrt{[(\lambda_1 + \lambda_2 + \lambda_3)^2 - 4(\gamma_1 + \gamma_2 + \gamma_3)]}, \tag{3.4}$$

with $\lambda_i = -\lambda_{ii}$ ($i = 1,2,3$), $\gamma_1 = \lambda_2\lambda_3 - \lambda_{23}\lambda_{32}$, $\gamma_2 = \lambda_1\lambda_3 - \lambda_{13}\lambda_{31}$ and $\gamma_3 = \lambda_1\lambda_2 - \lambda_{12}\lambda_{21}$. The explicit forms of the elements of $\mathbf{P}(t) = [P_{ji}(t)]$ are then given by (Chiang 1980, pp. 416-426):

$$P_{11}(t) = \gamma_1/(\alpha\beta) + \frac{\gamma_1 - \alpha(\beta - \lambda_1)}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_1 - \beta(\alpha - \lambda_1)}{\beta(\beta - \alpha)} \exp(-\beta t),$$

$$P_{21}(t) = \gamma_2/(\alpha\beta) + \frac{\gamma_2 - \alpha\lambda_{21}}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_2 - \beta\lambda_{21}}{\beta(\beta - \alpha)} \exp(-\beta t),$$

$$P_{12}(t) = \gamma_1/(\alpha\beta) + \frac{\gamma_1 - \alpha\lambda_{12}}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_1 - \beta\lambda_{12}}{\beta(\alpha - \beta)} \exp(-\beta t),$$

$$P_{22}(t) = \gamma_2/(\alpha\beta) + \frac{\gamma_2 - \alpha(\beta - \lambda_2)}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_2 - \beta(\alpha - \lambda_2)}{\beta(\beta - \alpha)} \exp(-\beta t),$$

(3.5)

$$P_{13}(t) = \gamma_1/(\alpha\beta) + \frac{\gamma_1 - \alpha\lambda_{13}}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_1 - \beta\lambda_{13}}{\beta(\beta - \alpha)} \exp(-\beta t),$$

$$P_{23}(t) = \gamma_2/(\alpha\beta) + \frac{\gamma_2 - \alpha\lambda_{23}}{\alpha(\alpha - \beta)} \exp(-\alpha t) + \frac{\gamma_2 - \beta\lambda_{23}}{\beta(\beta - \alpha)} \exp(-\beta t), \text{ and}$$

$$P_{3i}(t) = 1 - P_{1i}(t) - P_{2i}(t), \quad i = 1,2,3.$$

It is noted that

$$\lim_{t\downarrow 0} P_{ji}(t) = \begin{cases} 1, i = j \\ 0, \text{ otherwise,} \end{cases} \tag{3.6}$$

and that for all i,

$$\lim_{t\uparrow\infty} P_{ji}(t) = \gamma_j/(\alpha\beta) = \pi_j, (j = 1,2,3), \text{ say.} \tag{3.7}$$

The π_js, known as the (asymptotic) state probabilities, are independent of the initial state i. It is further noted that the elements of the rate matrix **K** contain structural information about the process, for example,

$$\lambda_i^{-1} = \text{ expected length of time (or mean residence time) for} \atop \text{a patient in state i to remain in that state.} \tag{3.8}$$

Also useful for checking the validity of parameter estimates (see section titled "Analysis of the ARC 090 Study and Conclusion") are the relations

$$\lambda_1 + \lambda_2 + \lambda_3 = \alpha + \beta \tag{3.9}$$

and

$$\gamma_1 + \gamma_2 + \gamma_3 = \alpha\beta, \tag{3.10}$$

which follow immediately from expression 3.4.

WEIGHTED CONDITIONAL NONLINEAR LEAST-SQUARES ESTIMATION

Suppose that the Markov process {X(t)} is piecewise stationary (Faddy 1976) so that the transition rates $\mathbf{K}(t) = [\lambda_{ji}(t)]$ may take on different sets of constant values for disjoint segments (time intervals) of {X(t)}. These stationary segments are in fact chosen to approximate the true process, which may be time dependent. The chosen segments should each contain a sufficient

number of observations to make the parameters (namely, the transition rates) statistically estimable. To fix the idea, let

$$\mathbf{K}(t) = \mathbf{K}_h \text{ for } \tau_{h-1} < t \le \tau_h, \, h = 1,2,..., \text{ and } \tau_0 = 0 \tag{4.1}$$

so that for each time interval $\tau_{h-1} < t \le \tau_h$, a unique solution $\mathbf{P}(t - \tau_{h-1})$ to equation 3.3 can be obtained in the same fashion that led to the explicit form given by equation 3.5. Suppose that within each such interval, observations are made at times $t_0 = \tau_{h-1} < t_1 < ... < t_m = \tau_h$. Let there be $N_j^{(h-1)}(\tau_{h-1})$ patients in state i (i = 1,2,3) at the end of the preceding segment, and for the sake of notational convenience, this number will be used interchangeably with $N_i^{(h)}(t_0)$. For now, let h be suppressed so that it will not be used to index the related expressions in the subsequent discussion. Furthermore, let $u_k = t_k - t_{k-1}$ k = 1,...,m, and suppose that the data consist of the number $X_{ji}(u_k)$ of patients who occupy state i at time t_{k-1} and state j at time t_k, i = 1,2,3, j = 1,2,3, k = 1,...m, where, by abuse of notation, the total number m of observations in a particular segment may not be the same as in the other. Then, given $N_j(t_{k-1})$ for i = 1,2,3, the component variables in each of the vectors

$$X_j(u_k) = [x_{1i}(u_k) x_{2i}(u_k) x_{3i}(u_k)] \tag{4.2}$$

where "'" means transposition, will follow a trinomial distribution with parameters $N_j(t_{k-1})$, $P_{1i}(u_k)$, $P_{2i}(u_k)$ and $P_{3i}(u_k) = 1 - P_{1i}(u_k) - P_{2i}(u_k)$. Let us also write

$$\mathbf{P}_i(u_k) = [P_{1i}(u_k), P_{2i}(u_k), P_{3i}(u_k)]' \tag{4.3}$$

and

$$\mathbf{Y}_i(u_k) = [X_{1i}(u_k), X_{2i}(u_k), X_{3i}(u_k)]'/N_i(t_{k-1}). \tag{4.4}$$

For simplicity, let us drop the argument u_k momentarily. It is thus easily seen that, for each i (i = 1,2,3),

$$E(\mathbf{Y}_i) = \mathbf{P}_i,$$

and

$$\text{Var}(\mathbf{Y}_i) = \frac{1}{N_i} \begin{bmatrix} P_{1i}(1-P_{1i}) & & \\ & & \text{Symm} \\ -P_{2i}P_{1i} & P_{2i}(1-P_{2i}) & \\ -P_{3i}P_{1i} & -P_{3i}P_{2i} & P_{3i}(1-P_{3i}) \end{bmatrix}$$

$$= \mathbf{\Phi}_i, \text{ say.}$$

142

At this point, it should be noted that the goal for this section is to set up a statistical model to estimate the parameters in each of the stationary pieces (or segments) of the Markov process {X(t)} represented by the parameter vector

$$[\lambda_{21}, \lambda_{31}, \lambda_{12}, \lambda_{32}, \lambda_{13}, \lambda_{23}, \alpha, \beta]', \qquad (4.5)$$

which will sometimes be written in an alternative expression as

$$\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8]'. \qquad (4.5a)$$

Because in the above equation the variance-covariance matrix Φ_i of Y_i is singular with rank equal to 2, we may proceed to estimate θ by utilizing just the 2x2 principal minor Ψ_i of Φ_i, along with the 2-vectors P_i^* and Y_i^* obtained by deleting the third elements of (4.3) and (4.4), respectively (i = 1,2,3). Next, let us string together these two sets of 2-vectors separately to form the following 6-vectors:

$$Q = [P_1^{*'}, P_2^{*'}, P_3^{*'}]'$$

and

a $$Z = [Y_1^{*'}, Y_2^{*'}, Y_3^{*'}]'.$$

It then becomes obvious that

$$E(Z) = Q$$

and

$$Var(Z) = Diag[\Psi_1, \Psi_2, \Psi_3] = \Psi, \text{ say.}$$

By replacing the argument u_k, we can now write a statistical model as follows:

$$Z(u_,) = Q(u_k) + e(u_k), \ k = 1,...,m, \qquad (4.6)$$

where the elements of $Q(u_k)$ are nonlinear functions of θ, $E\{e(u_k)\} = 0$, and $Var\{e(u_,)\} = \Psi(u_k)$ for all k. It remains to find the least-squares estimates for the parameters Θ by fitting the data from the ARC 090 study, prearranged in the

form (4.4), to the nonlinear model (4.6). Before taking this step, let us make a linear transformation of model (4.6) to remove the statistical dependence in its error structure. Note that each of the diagonal blocks $\Psi_{:}(u_k)$ in $\Psi((u_k)$ has the following   expression:

$$\Psi_i = \begin{bmatrix} P_{1i}(1-P_{1i}) & -P_{1i}P_{2i} \\ -P_{2i}P_{1i} & P_{2i}(1-P_{2i}) \end{bmatrix} , \; i = 1,2,3,$$

where the argument u_k is again suppressed. Because the variance-covariance matrix $\Psi((u_k)$ is positive definite, there exists a nonsingular lower triangular matrix $T(u_k)$ such that

$$\Psi(u_k) = T(u_k)T'(u_k).$$

Specifically, $T(u_k) = \text{Diag}[T_1(u_k), \; T_2(u_k), \; T_3(u_k)]$ with

$$T_I = \begin{bmatrix} \sqrt{[P_{1i}(1-P_{1i})]} & 0 \\ -P_{2i}\sqrt{P_{1i}}/\sqrt{(1-P_{1i}} & \sqrt{(P_{2i}P_{3i})}/\sqrt{(1-P_{1i})} \end{bmatrix} i = 1,2,3,$$

the argument u_k having been omitted. By virtue of the matrix $T(u_k)$, we can then deal with the new transformed model

$$T^1(u_k)Z(u_k) = T^1(u_k)Q(u_k) + e(u_k), \; k = 1,...,m, \qquad (4.7)$$

where $\varepsilon(u_k) = T^{-1}(u_k)e(u_k)$ is the new error term with $E\varepsilon(u_k) = 0$ and $Var\varepsilon(u_k) = I$, a 6x6 identity matrix. Thus, we see that the new model (4.7) has achieved an independent error structure. Based on this model, we can now find the least-squares estimate for θ by minimizing the object function

$$S(\theta) = \Sigma_k\varepsilon'(u_k)\varepsilon(u_k)$$

$$\qquad (4.8)$$

$$= \Sigma_k[Z(u_k)-Q(u_k)]'\Psi^{-1}(u_k)[Z(u_k)-Q(u_k)].$$

This may be achieved by an iterative SAS function minimization program (SAS Institute, Inc. 1988, chapter 23) using the derivative-free DUD option (Ralston and Jennrich 1978). The starting values for θ can be derived from the maximum likelihood estimates (MLEs) for the transition probabilities P_{ji} associated with a discrete-time Markov chain $\{X(k): k = 1,2,...\}$ resulting from the original continuous-time Markov process $\{X(t): 0<t<\infty\}$ when periodic observations are made on the latter. Specifically, let $X_{ji}(k)$ be the number of transitions $i{\rightarrow}j$ during the interval $u_k = t_k - t_{k-1}$, $k = 1,2,...m$, where m is the total number of observations made within a particular stationary segment of $\{X(t)\}$. Then, given the initial state i (i=1,2,3), the joint trinomial distribution of $X_{ji}(k)$, aggregated over k, will produce the MLEs of P_{ji} in the form

$$\hat{P}_{ji} = \Sigma_k X_{ji}(k)/[\Sigma_k N_i(k-1)] \tag{4.9}$$

where $N_i(k - 1) = \Sigma_j X_{ji}(k)$ is the number of patients in state i at time t_{k-1}, $k = 1,2,...,m$. If we denote by u the length of the interval between successive observation times, then it follows from the first equation of (3.2) that

$$\lambda_{ji} \approx P_{ji}/u, \ j{\neq}j \tag{4.10}$$

Because, according to the study design, observations were taken every other day except for the weekend, we can roughly set u = 2 and use the quotient $\lambda^0_{ji} := P_{ji}/2$ as the initial estimate for λ_{ji}, $j{\neq}i$ (i,j = 1,2,3). These values are in turn substituted in equation (3.4) to obtain the initial estimates for a and ß.

ANALYSIS OF THE ARC 090 STUDY AND CONCLUSION

Some details on this clinical trial have been given in the introduction. For the purpose of saving space, *only* a subset of the observed data is presented here (table 1). The first column gives the treatment indicator TRT, which has value 1 if the treatment was buprenorphine, 2 if it was methadone 20 mg, and 3 if it was methadone 60 mg. The second and third columns give, respectively, the serial observation times t (three times per week taken on Monday, Wednesday, and Friday) and the intervals u in days between successive observation times. These are followed by three columns of state totals N_i (i = 1,2,3) representing the total numbers of patients in state i at various observation times. Because the state total N_3 is of composite nature, it is further split into two components, M and L, which occupy the next two columns, with M and L representing, respectively, the number of patients who missed the urinary test and the number of patients who were lost to followup. The last nine columns are

TABLE 1. *Partial listing of observed data from ARC 090 study*

TRT	t	U	Stale Total			N₃	Transition Count								
			N_1	N_2	N_3	(M, L)	X_{11}	X_{21}	X_{31}	X_{12}	X_{22}	X_{32}	X_{13}	X_{23}	X_{33}
	0	0	7	36	10	(10.0)									
	1	2	17	30	6	(5.1)	5	2	0	7	24	5	5	4	1
	2	2	27	18	8	(6, 2)	14	2	1	11	14	5	2	2	2
	3	3	23	19	11	(8, 3)	20	5	2	3	11	4	0	3	5
	4	2	22	18	13	(7, 6)	17	3	3	3	14	2	2	1	8
	5	2	26	13	12	(6, 6)	21	0	1	4	10	4	3	3	7
	6	3	26	16	11	(5, 6)	24	4	0	2	9	2	0	3	9
	7	2	21	16	16	(7, 9)	20	4	2	1	10	5	0	2	9
	8	2	23	13	17	(7, 10)	18	2	1	5	7	4	0	4	12
	9	3	28	12	13	(3, 10)	20	2	1	6	7	0	2	3	12
	10	2	25	12	16	(5, 11)	23	3	2	2	8	2	0	1	12
	11	2	26	10	17	(6, 11)	22	2	1	2	7	3	2	1	13
	12	3					20	2	4	4	6	0	1	2	14
SUM			294	189	153		224	31	18	50	127	36	17	29	104
MLE							.82	.11	.07	.24	.60	.17	.11	.19	.69
2	0	0	2	46	7	(7, 0)									
2	1	2	18	29	8	(8, 0)	2	0	0	15	23	8	1	6	0
2	2	2	16	31	8	(8, 0)	10	5	3	5	23	1	1	3	4
2	3	3	17	28	10	(8, 2)	14	1	1	2	24	5	1	3	4
2	4	2	15	25	15	(11, 4)	12	3	2	1	21	6	2	1	7
2	5	2	15	21	19	(14, 5)	8	5	2	5	16	4	2	0	13
2	6	3	19	20	16	(8, 8)	9	5	1	6	13	2	4	2	13
2	7	2	6	25	24	(14, 10)	3	7	9	2	13	5	1	5	10
2	8	2	10	25	20	(8, 12)	3	3	0	4	18	3	3	4	17
2	9	3					7	3	0	5	15	5	2	1	17
SUM			118	250	127		68	32	18	45	166	39	17	25	85
MLE							.58	.27	.15	.18	.66	16	.13	.20	.67
3	0	0	6	43	5	(5, 0)									
3	1	2	16	32	6	(6, 0)	5	2	0	10	28	5	1	3	1
3	2	2	22	23	9	(7, 2)	11	2	3	9	19	4	2	2	2
3	3	3	17	23	14	(10, 4)	15	4	3	1	16	6	1	3	5

TABLE 1. *(continued)*

TRT	t	U	State Total			N_3	Transition Count								
			N_1	N_2	N_3	(M, L)	x_{11}	x_{21}	x_{31}	x_{12}	x_{22}	x_{32}	x_{13}	x_{23}	x_{33}
3	4	2	16	23	15	(11, 4)	9	6	2	5	14	4	2	3	9
3	5	2	18	20	16	(11, 5)	10	3	3	7	13	3	1	4	10
3	6	3	14	22	18	(7, 11)	11	5	2	2	13	5	1	4	11
3	7	2	18	20	16	(3, 13)	10	2	2	7	14	1	1	4	13
3	8	2	16	21	17	(4, 13)	13	5	0	2	16	2	1	0	15
3	9	3					11	3	2	4	14	3	2	2	13
SUM			143	227	116		95	31	17	47	147	33	12	25	79
MLE							.66	.22	.12	.21	.65	.15	.10	.22	.68
1	36	3	14	6	4	(4.29)									
1	37	2	15	4	5	(5, 29)	13	0	1	1	3	2	1	1	2
1	38	2	13	7	4	(4, 29)	11	3	1	1	2	1	1	2	2
1	39	3	13	5	6	(5, 30)	9	2	2	4	2	1	0	1	3
1	40	2	12	7	5	(4, 30)	8	5	0	2	1	2	2	1	3
1	41	2	16	2	6	(5, 30)	11	0	1	4	2	1	1	0	4
1	42	3	13	3	8	(7, 30)	10	1	5	0	2	0	3	0	3
1	43	2	14	5	5	(4, 30)	10	1	2	2	1	0	2	3	3
1	44	2	15	6	3	(2, 30)	10	3	1	2	3	0	3	0	2
1	45	3	13	4	7	(6, 30)	10	3	2	3	1	2	0	0	3
1	46	2	14	7	3	(2, 30)	11	2	0	1	3	0	2	2	3
1	47	2	14	7	3	(2, 30)	11	2	1	3	4	0	0	1	2
1	46	3	14	2	8	(7, 30)	11	0	3	3	1	3	0	1	2
1	49	2	11	6	7	(6, 30)	9	3	2	1	0	1	1	3	4
1	50	2	14	7	3	(1, 31)	10	1	0	3	3	0	1	3	3
1	51	3					11	1	2	3	2	2	0	0	3
SUM			205	78	77		155	27	23	33	30	15	17	18	42
MLE							.76	.13	.11	.42	.39	.19	.22	.23	.55
2	42	3	3	8	1	(1, 43)									
2	43	2	7	4	1	(1, 43)	3	0	0	3	4	1	1	0	0
2	44	2	5	6	1	(1, 43)	3	4	0	2	2	0	0	0	1
2	45	3	5	6	1	(0, 44)	3	1	1	1	5	0	1	0	0

TABLE 1. *(continued)*

TAT	t	u	N₁	N₂	N₃	(M, L)	x_{11}	x_{21}	x_{31}	x_{12}	X_{22}	x_{32}	x_{13}	X_{23}	x_{33}
				State Total		N_3					Transition Count				
2	46	2	4	6	2	(1, 44)	4	0	1	0	6	0	0	0	1
2	47	2	3	8	1	(0, 44)	3	1	0	0	6	0	0	1	1
2	46	3	4	7	1	(0, 44)	2	1	0	2	6	0	0	0	1
2	49	2	3	7	2	(1, 44)	3	1	0	0	6	1	0	0	1
2	50	2	4	7	1	(0, 44)	3	0	0	1	6	0	0	1	1
2	51	3					3	1	0	0	6	1	0	0	1
SUM			38	59	11		27	9	2	9	47	3	2	2	7
MLE							.71	.24	.05	.15	.79	.05	.18	.18	.64
3	42	3	11	10	4	(4, 29)									
3	43	2	11	10	4	(3, 30)	9	1	1	2	6	0	0	1	3
3	44	2	13	7	5	(2, 32)	9	1	1	3	6	1	1	0	3
3	45	3	9	4	12	(8, 33)	8	0	5	1	4	2	0	0	5
3	46	2	7	6	10	(6, 33)	6	1	2	0	4	0	1	3	8
3	47	2	7	6	10	(6, 33)	5	1	1	2	5	1	0	2	8
3	48	3	7	8	10	(6, 33)	3	4	0	3	4	1	1	0	9
3	49	2	5	10	10	(5, 34)	2	4	1	2	5	1	1	1	8
3	50	2	7	5	13	(7.35)	5	0	0	2	4	4	0	1	9
3	51	3					6	1	0	2	3	0	0	2	11
SUM			77	70	78		51	15	11	17	43	10	4	10	64
MLE							.66	.20	.14	.24	.61	.14	.05	.13	.82

KEY: M = missing; L = lost; TRT = treatment; SUM = sum; MLE = maximum likelihood estimate

allocated to the transition counts X_{ji} (i,j = 1,2,3) representing the number of patients who were observed to undergo the transition i→j during an interval u. At the bottom of each segment the sums of N_i and X_{ji} are given, followed by the MLEs for transition probabilities calculated from equation (4.9).

To estimate the parameters in (4.5), the observed serial data were partitioned into disjoint stationary segments (3 to 6 weeks in length), which were empirically identified and fitted to the model (3.5). These segments were selected by trial and error to ensure the following: (1) a convergence for the weighted conditional nonlinear least-squares estimation procedure described in the last section-the convergence criterion used was "$S(\theta) < 10^{-8}$"; (2) a reasonable closeness between the estimate of $\sum_i \lambda_i$ (i.e., $\sum_i \sum_j \lambda_{ji}$, j ≠ i) and that of $\alpha + \beta$— see equation (3.9); and (3) a good fit of the data to the model. The data from treatments 1, 2, and 3 comprised four, five, and four such segments, respectively. Of these, the first and the last segments from each treatment are shown in table 1. It should be noted that, barring the first segment for each treatment, all subsequent segments contained data from only those patients who had stayed in the study long enough to enter into these segments, although some of the segment members might still fail to stay for the whole segment-for example, taking the last segment of treatment 1, the 29 (cumulative) losses at the end of the preceding segment were not counted, whereas the two losses (= 31 - 29) that occurred during this last segment remained nominally counted. Moreover, data from the week between the fourth and the fifth segments for treatment 2 as well as those from the week between the second and the third segments for treatment 3 were omitted because inclusion of them in either of the neighboring segments would have resulted in poor fit of the data to the model. Table 2 lists by treatment and segment the least-squares estimates of the transition rates $[\lambda_{ji}]$ and the transition probabilities $[P_{ji}(u)]$, where u = 2 days if transition counts were made every other day during weekdays and u = 3 days if counts were made over weekends.

As can be seen from the last block of columns in table 2, the observed data, after having been partitioned into empirically identified stationary segments, appear to accord well with the model. With estimates for the rate matrix and the transition matrix in place, some useful information may now be extracted from these estimates. Here the transition rates λ_{ji}, j ≠ i, measure on a daily basis the propensity for a patient to move between pairs of disease states, granting that

TABLE 2. *Least-squares estimates of transition rates and transition probabilities for empirically identified stationary segments, listed by treatment*

		Rate Matrix $K = [\lambda_{ij}]$			Transition Matrices $P = [P_{ij}(2)]$			$P = [P_{ij}(3)]$			Goodness of Fit DF	X^2	p
Seg	Wks												
						Treatment = Buprenorphine							
1	4	-.087	.102	.089	.857	.160	.098	.802	.216	.141	64	79.32	.0939
		.055	-.184	.054	.086	.718	.130	.117	.624	.171			
		.032	.082	-.143	.057	.122	.772	.081	.160	.689			

$\alpha = .283$ $\beta = .168$ $\alpha+\beta = .451$ $\sum\lambda_i = -\sum\lambda_{ii} = .413$

2	3	-.127	.135	.049	.799	.195	.095	.728	.254	.139	46	50.10	.3140
		.080	-.233	.114	.119	.663	.160	.157	.562	.204			
		.047	.098	-.163	.082	.141	.745	.115	.183	.657			

$\alpha = .341$ $\beta = .209$ $\alpha+\beta = .550$ $\sum\lambda_i = -\sum\lambda_{ii} = .523$

3	5	-.114	.124	.044	.809	.190	.086	.738	.252	.125	82	82.22	.4724
		.070	-.219	.071	.108	.662	.110	.144	.551	.146			
		.044	.095	-.116	.082	.148	.804	.118	.198	.729			

$\alpha = .218$ $\beta = .209$ $\alpha+\beta = .427$ $\sum\lambda_i = -\sum\lambda_{ii} = .449$

4	5	-.122	.212	.111	.816	.294	.191	.756	.376	.264	82	71.39	.7923
		.066	-.308	.117	.096	.577	.148	.125	.461	.182			
		.056	.096	-.228	.088	.129	.661	.119	.163	.554			

$\alpha = .392$ $\beta = .292$ $\alpha+\beta = .684$ $\sum\lambda_i = -\sum\lambda_{ii} = .658$

TABLE 2. *(continued)*

		Rate Matrix			Transition Matrices						Goodness of Fit		
Seg	Wks	$K = [\lambda_{ij}]$			$P = [P_{ij}(2)]$			$P = [P_{ji}(3)]$			DF	X^2	p
					Treatment = Methadone 20 mg								
1	3	-.214	.088	.067	.671	.131	.106	.565	.171	.143	46	49.99	.3179
		.139	-.164	.096	.207	.746	.156	.270	.663	.211			
		.075	.076	-.163	.122	.123	.738	.165	.166	.646			

$\alpha = .277 \quad \beta = .239 \quad \alpha+\beta = .516 \quad \Sigma\lambda_i = -\Sigma\lambda_{ii} = .541$

2	4	-.218	.080	.020	.675	.111	.037	.571	.141	.054	64	68.41	.3300
		.129	-.143	.079	.186	.782	.125	.242	.708	.170			
		.089	.063	-.099	.139	.107	.838	.187	.150	.776			

$\alpha = .324 \quad \beta = .194 \quad \alpha+\beta = .515 \quad \Sigma\lambda_i = -\Sigma\lambda_{ii} = .459$

					Treatment = Buprenorphine								
3	3	-.188	.092	.027	.711	.133	.052	.615	.172	.076	46	38.09	.7901
		.152	-.165	.087	.219	.751	.138	.283	.671	.187			
		.035	.073	-.114	.070	.116	.810	.102	.157	.737			

$\alpha = .301 \quad \beta = .183 \quad \alpha+\beta = .484 \quad \Sigma\lambda_i = -\Sigma\lambda_{ii} = .467$

4	3	-.209	.110	.031	.688	.151	.062	.589	.192	.090	46	61.55	.0623
		.164	-.202	.123	.228	.712	.181	.290	.628	.237			
		.045	.092	-.154	.084	.137	.758	.121	.180	.673			

$\alpha = .353 \quad \beta = .231 \quad \alpha+\beta = .584 \quad \Sigma\lambda_i = -\Sigma\lambda_{ii} = .565$

5	3	-.145	.077	.091	.765	.125	.144	.681	.170	.193	46	37.57	.8074
		.119	-.102	.100	.192	.833	.168	.260	.773	.231			
		.026	.025	-.191	.043	.042	.688	.058	.057	.576			

$\alpha = .222 \quad \beta = .207 \quad \alpha+\beta = .429 \quad \Sigma\lambda_i = -\Sigma\lambda_{ii} = .438$

TABLE 2. *(continued)*

		Rate Matrix $K = [\lambda_{ji}]$			Transition Matrices $P = [P_{ji}(2)]$			$P = [P_{ji}(3)]$			Goodness of Fit DF	X^2	p
Seg	Wks												
						Treatment = Methadone 60 mg							
1	3	-.169	.104	.052	.743	.150	.089	.657	.196	.125	46	48.36	.3778
		.109	-.177	.108	.161	.739	.160	.212	.656	.211			
		.060	.073	-.160	.096	.111	.751	.131	.148	.664			
		$\alpha = .283$ $\beta = .168$ $\alpha+\beta = .451$ $\Sigma\lambda_i = -\Sigma\lambda_{ii} = .413$											
2	4	-.151	.109	.043	.778	.167	.079	.700	.218	.113	64	53.49	.8227
		.094	-.184	.087	.131	.724	.131	.174	.635	.174			
		.057	.075	-.130	.091	.109	.790	.126	.147	.713			
		$\alpha = .281$ $\beta = .232$ $\alpha+\beta = .513$ $\Sigma\lambda_i = -\Sigma\lambda_{ii} = .495$											
3	6	-.101	.115	.050	.763	.162	.093	.681	.212	.131	100	79.63	.9337
		.071	-.179	.097	.144	.722	.160	.191	.633	.210			
		.030	.064	-.147	.093	.116	.747	.128	.155	.659			
		$\alpha = .274$ $\beta = .189$ $\alpha+\beta = .463$ $\Sigma\lambda_i = -\Sigma\lambda_{ii} = .427$											
4	3	-.169	.122	.026	.735	.173	.051	.644	.221	.076	46	41.81	.6483
		.098	-.194	.064	.143	.706	.101	.187	.610	.135			
		.072	.072	-.090	.122	.121	.848	.169	.169	.788			

$\alpha = .291$ $\beta = .176$ $\alpha+\beta = .467$ $\Sigma\lambda_i = -\Sigma\lambda_{ii} = .452$

transition does take place, and these rates are in turn related to $P_{ji}(u)$, u = 2 or 3, by equations (3.2) and (4.10). Now let us proceed with comparison of treatment effects by looking into the difference $\lambda_{12} - \lambda_{21}$ over successive segments, as given below:

Segments:	1	2	3	4	5
Buprenorphine:	0.047	0.055	0.054	0.156	
Methadone 20 mg:	-0.051	-0.049	-0.060	-0.054	-0.042
Methadone 60 mg:	-0.005	0.015	0.044	0.024	

From these comparisons it is clear that, in contrast to both buprenorphine and methadone 60 mg, methadone 20 mg tended to induce a greater rate of transition from state 1 to state 2 than from state 2 to state 1 (which refers to an improvement in the clinical sense); however, buprenorphine is superior even to methadone 60 mg because, from the first segment on, it reversed the direction of transition in its favor, with its strength culminating in an impulsive surge at the last segment. This tendency appeared to be unperturbed by the continual loss of patients during the course of the study because (1) between-group differences were seen to be highly consistent from the first segment (which had kept all patients formally under observation) till the last segment (which left out all patients who had been lost to followup in the preceding segments), and (2) both the buprenorphine group and the methadone 60 mg group had about the same patient-loss rate (63.0 percent for the latter vs. 60.4 percent for the former according to Jain, "Analysis of Clinical Trials," this volume) during the course of the study. However, the substantially increased loss rate (80.0 percent) seen in the methadone 20 mg group could mostly be ascribed to its poor performance.

The above observation is also applicable to $P_{ji}(u)$, which refers to the likelihood of the transition i→j (j ≠ i) during an interval u. This likelihood tended to be greater with u = 3 days than with u = 2 days, irrespective of the direction of transition. That is, more transitions are likely to occur during weekends than during weekdays.

Put in another perspective, the effect of treatment may also be reflected by the duration of improvement defined as the mean length of time a patient spends in state 1 during each segment-see expression (3.8) and table 3. It was found out that, once the patient had been in state 1 (i.e., negative for opiate), he or she stayed there much longer when on buprenorphine (7.9 to 11.5 days) than when on methadone 20 mg (4.7 to 6.9 days), with his or her duration of improvement ranked in between when on methadone 60.mg (5.9 to 9.9 days). The same was also true when the effect of treatment was evaluated in terms of the asymptotic state probabilities for the successive segments-see equation (3.7) and table 3. These state probabilities may be interpreted as the expected state-specific proportions of patients in the population, assuming that each

TABLE 3. *Mean residence time and asymptotic state probabilities*

Seg	Wks	Mean Residence Time			Asymptotic State Probability		
		St 1	St 2	St 3	π_1	π_2	π_3
			Treatment = Buprenorphine				
1	4	11.5	5.4	7.0	.474	.266	.260
2	3	7.9	4.3	6.1	.419	.288	.293
3	5	8.8	4.6	8.6	.403	.244	.353
4	5	8.2	3.3	4.4	.566	.208	.226
			Treatment = Methadone 20 mg				
1	3	4.7	6.1	6.2	.268	.414	.318
2	4	4.6	7.0	10.2	.183	.397	.420
3	3	5.3	6.0	8.8	.249	.409	.342
4	3	4.8	5.0	6.5	.264	.412	.324
5	3	6.9	9.8	5.2	.353	.527	.119
			Treatment = Methadone 60 mg				
1	3	5.9	5.7	6.3	.325	.381	.294
2	4	6.6	5.4	6.3	.354	.352	.293
3	6	9.9	5.6	6.8	.465	.308	.227
4	3	5.9	5.2	11.2	.271	.285	.444

segment continues to maintain its own stochastic characteristics beyond the segmental cutoff point.

Based on the results of this section, we may draw the following conclusion: Relative to methadone, buprenorphine, as demonstrated in a randomized comparative clinical trial, is found to be a substantially effective drug indicated for the treatment of opiate addiction with steadfast action-patients on buprenorphine treatment tend to achieve an opiate-negative state faster, and on reaching that state, they stay there longer.

DISCUSSION

The stochastic compartmental model proposed in this chapter provides a probabilistic description of disease-state transitions for opiate-addicted patients under medical treatment whose disease states were monitored by serially administered urinary tests. The result of a test at each visit was retrospectively assigned to one of three compartments numbered 1, 2, and 3 according to

whether it was (1) negative for opiate, (2) positive for opiate, or (3) test was missing or patient was lost to followup. The dynamics of changes in the disease state were then analyzed by a regular (all transition probabilities being nonzero) three-state continuous-time Markov process with constant transition rates. The author is aware of the fact that lumping patients who missed the urinary test and those who were lost to followup in a composite compartment, numbered 3 in the above, would result in underestimated rates of transition out of that compartment. Therefore, these rates were deliberately slighted in the reporting of results because they were not as important as the remaining ones, which still contain unabridged information. In theory, it is known that an arbitrary rearrangement of the state space associated with a Markov chain may render the new chain non-Markovian (Kingman 1972; Kemeny and Snell 1960). In applications, however, all that is required is that the model employed be sufficiently near to reality. Therefore, rearrangements are often made on states regarded as less germane to the question of interest (Bartholomew 1982; Fix and Neyman 1951). The three criteria, (1) through (3), used in the preceding section for identification of stationary segments as well as the partial truncation of lost patients from successive segments are in fact the embodiment of practical efforts made to ensure that the theoretically coarser model employed in the analysis is sufficiently correct. Naturally, we could have separated the two states M and L in compartment 3 to model the data via an absorbing Markov chain or, perhaps, via a semi-Markov process. However, this would not only introduce more parameters to be estimated but also shift the focus of the problem toward examining the clinically less important parameter, namely, the mean absorption time. The author is nevertheless still intent to pursue the problem along these lines as a supplementary analysis of the ARC 090 study.

Originally, the ARC 090 data set did contain values of several prognostic variables such as the age and sex of the patient. However, the possible effects of these variables were not taken into consideration in the model employed above. Based on the fact that the ARC 090 trial was a genuine prospective study having been executed under strict adherence to the principle of experimental design and randomization, it is deemed unlikely that the treatment comparisons had been severely confounded by these covariates.

It is also perhaps proper and necessary to add that the SAS program for function minimization utilized to estimate the rates of disease-state transition was capable of providing neither a valid asymptotic standard error for each of the estimates nor an asymptotic correlation matrix for them. This limitation has precluded the present approach to the analysis of the ARC 090 study from going into the realms of interval inferences and/or hypothesis testings. It is hoped that this deficiency will be resolved subsequently.

REFERENCES

Bartholomew, D.J. *Stochastic Models for Social Processes.* New York: Wiley, 1982.

Chiang, C.L. *An Introduction to Stochastic Processes and Their Applications.* New York: Krieger, 1980.

Faddy, M.J. A note on the general time-dependent stochastic compartmental model. *Biometrics* 32:443-448, 1976.

Fix, E., and Neyman, J. A simple stochastic model of recovery, relapse, death, and loss of patients. *Hum Biol* 23:205-241, 1951.

Gail, M.H. Evaluating serial cancer marker studies in patients at risk of recurrent disease. *Biometrics* 37:67-78, 1981.

Gehan, E.A. A generalized two-sample Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52:203-223, 1965.

Kemeny, J.G., and Snell, J.L. *Finite Markov Chains.* Princeton, NJ: Van Nostrand, 1960.

Kingman, J.F.G. *Regenerative Phenomena.* New York: Wiley, 1972.

Lin, D.Y. MULCOX: A computer program for the Cox regression analysis of multiple time variables. *Comput Methods Programs Biomed* 32:125-135, 1990.

Pete, R., and Peto, J. Asymptotically efficient rank invariant test procedure (with discussion). *J Roy Stat Society, Ser. A* 135:185-206, 1972.

Prentice, R.L.; Williams, B.J.; and Peterson, A.V. On the regression analysis of multivariate failure time data. *Biometrika* 68:373-379, 1981.

Ralston, M.L., and Jennrich, R.I. DUD, a derivative-free algorithm for nonlinear least squares. *Technometrics* 20:7-14, 1978.

SAS Institute, Inc. *SAS/STAT User's Guide, Release 6.03 Edition.* Cary, NC: SAS Institute, Inc., 1988.

Wei, L.J., and Lachin, J.M. Two-sample asymptotically distribution-free tests for incomplete multivariate observations. *J Am Stat Assoc* 79:653-661, 1984.

Wei, L.J.; Lin, D.Y.; and Weissfeld, L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc* 84:1065-1073, 1989.

ACKNOWLEDGMENT

AUTHOR

T.S. Weng, Ph.D.
Mathematical Statistician

Division of Biometric Sciences, HFZ-162
Office of Science and Technology
Center for Devices and Radiological Health
Food and Drug Administration
Suite 405
1801 Rockville Pike
Rockville, MD 20852

Summary of Discussion: "Toward a Dynamic Analysis of Disease-State Transition Monitored by Serial Clinical Laboratory Tests" by Weng*

Alan J. Gross

In his chapter, Dr. Weng considers a three-compartment model applicable to data generated from a randomized clinical trial designed to evaluate the efficacy of buprenorphine for the treatment of opiate addiction. The model allows for missing observations due to schedule noncompliance or lost-to-followup patients.

Although the transition probabilities are expected to be time dependent, the time axis is subdivided into intervals over which the transition rates are assumed to be constant.

As indicated during the discussion, several generalizations of this model are possible:

- A four-compartment model can be considered whereby those patients who are lost to followup or censored are in a compartment separate from the noncompliant patients. Since such patients always remain in this compartment once they have entered it, it is an absorbing compartment or state.

- A mover-stayer model was also suggested whereby a fraction of the patients would always remain in a state whenever they were in it. This would allow a fraction of addicts to be considered "cured" and another fraction to be considered "not cured" in spite of the intervention.

*The chapter by Dr. Weng was reviewed by Dr. Greenhouse prior to the technical review. Dr. Gross reviewed the revised chapter submitted by Dr. Weng and wrote this summary of the discussion that took place at the meeting.

- The same transition rates were not assumed in the distinct treatment groups. In fact, methodology may be developed to test the hypothesis that the transition rates are equal across the treatment groups, leading to development of statistical inference techniques for this type of model.

The difficulties that need to be addressed in using this model include (1) determining the points on the time axis where the parameters change and (2) how the parameters are to be estimated given the data.

Overall, the model is an excellent beginning for the analysis of this type of clinical trial data and with modification can be very useful.

AUTHOR

Alan J. Gross, Ph.D.
Professor
Department of Biostatistics, Epidemiology, and Systems Science
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2503

A Markov Model for NIDA Data on Treatment of Opiate Dependence*

Mei-Ling Ting Lee

INTRODUCTION

This chapter models the data set collected from the first 17 weeks of the National Institute on Drug Abuse ARC 090 trial by two-state Markov chains.

For complete data, empirical Bayes procedures are considered. Assuming that the one-step transition probabilities have independent Beta priors, the posterior transition probabilities are derived. As a consequence, the expected number of positive results in a sequence of $m + 1$ urine samples can be estimated. For cases where data sets contain missing values, we consider the adjusted likelihood function obtained by taking into account the k-step transition probabilities contributed by $(k - 1)$ consecutive missing values and compute the maximum likelihood estimates (MLEs) for the one-step transition matrix by the Newton-Raphson method.

COMPLETE DATA

A Two-State Markov Chain

Assume that the sequence of a patient's urine sample test results, denoted by $X = (X_l, \ldots, X_{m+1})$, forms a sequence of Markov-dependent Bernoulli trials, where X_i takes on values of 0 or 1, representing negative or positive results, respectively.

Assume that the process is time homogeneous; that is, transition probabilities

$$p_{ij} = P(X_{l+1} = j \mid X_l = i), \; i, j \in \{0, 1\},$$

are independent of l. Also, assume that $0 < p_{ij} < 1$ for all $i, j \in \{0, 1\}$. Then, by the Markov property, the likelihood function based on the sample sequence

---

*Computer programs used to do the calculations in this chapter are available upon request.

of test results of a patient is given by

$$
\begin{aligned}
f(x_1, &\ldots, x_{m+1} \mid p_{01}, p_{10}) \\
&= P(X_1 = x_1) P(X_2 = x_2 \mid X_1 = x_1) \cdots P(X_{m+1} = x_{m+1} \mid X_m = x_m) \\
&= P(X_1 = x_1) p_{00}^{n_{00}} p_{01}^{n_{01}} p_{10}^{n_{10}} p_{11}^{n_{11}} \\
&= P(X_1 = x_1)(1 - p_{01})^{n_{00}} p_{01}^{n_{01}} p_{10}^{n_{10}} (1 - p_{10})^{n_{11}}
\end{aligned}
\tag{1}
$$

where n_{ij} = { the number of times $X_{l-1} = i$ is followed by $X_l = j$, $l = 2, \ldots, m + 1$ } is the one-step transition count.

Let $p_0^{(k)} = P(X_k = 0)$ and $p_1^{(k)} = P(X_k = 1)$ denote the state probabilities of the kth test result of a patient. Let $p_0 = \lim_{k \to \infty}$ and $p_1 = \lim_{k \to \infty} p_1^{(k)}$ denote the stationary probabilities of the urine sample testing process of this patient when the process has achieved statistical equilibrium after a large number of tests have been taken. Then we have

$$
(p_0, p_1) = (p_{10}/(p_{10} + p_{01}), \; p_{01}/(p_{10} + p_{01})).
\tag{2}
$$

If tests are given systematically with an equal number of days between tests, then the recurrence time can be estimated as follows. Suppose a patient's urine sample result is positive at the first trial. Let t denote the recurrence time, that is, the time of his or her next positive result. That is, $X_1 = 1$, $X_2 = \ldots = X_{t-1} = 0$, and $X_t = 1$. Let T be the random variable of the recurrence time, then

$$
E(T) = (p_{01} + p_{10})/p_{01}.
\tag{3}
$$

Moreover, let $Y_{m+1} = \sum_{i=1}^{m+1} X_i$ denote the number of positive results in a sequence of $m + 1$ urine samples from a patient. Then Y_{m+1} is asymptotically normal with mean

$$
E(Y_{m+1}) \sim (m + 1) p_{01}/(p_{01} + p_{10})
$$

and variance

$$
\text{Var}(Y_{m+1}) \sim (m + 1) p_{01} p_{10} (2 - p_{01} - p_{10})/(p_{01} + p_{10})^3.
\tag{4}
$$

See Cox and Miller (1965) for a review of properties of Markov chains.

An Empirical Bayes Approach

Bayes decision methods for finite-state Markov chains with complete observations have been discussed by Martin (1967) and Basawa and Prakasa Rao (1980). We consider in this section empirical Bayes procedures.

If results for sequences of $m + 1$ urine samples are collected from each of the r patients, chances are that transition probabilities p_{01} and p_{10} will vary from patient to patient. Let U and V denote random variables that correspond to the respective transition probabilities p_{01} and p_{10}, which vary across patients. For simplicity, equation (1) can be rewritten as the following form:

$$f(x_1, \ldots, x_{m+1} \mid u, v) = P(X_1 = x_1)(1 - u)^{n_{00}} u^{n_{01}} v^{n_{10}} (1 - v)^{n_{11}} \qquad (1a)$$

Assume that U and V are independent with prior distributions $\mathrm{Beta}(\alpha_1, \beta_1)$ and $\mathrm{Beta}(\alpha_2, \beta_2)$, respectively. Let g_1, g_2 denote their corresponding density functions. Then we have $g(u, v \mid \alpha_1, \beta_1, \alpha_2, \beta_2) = g_1(u \mid \alpha_1, \beta_1) g_2(v \mid \alpha_2, \beta_2)$ where

$$g_1(u \mid \alpha_1, \beta_1) = \frac{\Gamma(\alpha_1 + \beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)} u^{\alpha_1 - 1}(1 - u)^{\beta_1 - 1} = \frac{1}{B(\alpha_1, \beta_1)} u^{\alpha_1 - 1}(1 - u)^{\beta_1 - 1}$$

$$g_2(v \mid \alpha_2, \beta_2) = \frac{\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_2)\Gamma(\beta_2)} v^{\alpha_2 - 1}(1 - v)^{\beta_2 - 1} = \frac{1}{B(\alpha_2, \beta_2)} v^{\alpha_2 - 1}(1 - v)^{\beta_2 - 1}$$

Hence,

$$f(x_1, \ldots, x_{m+1} \mid \alpha_1, \beta_1, \alpha_2, \beta_2)$$

$$= \int_0^1 \int_0^1 f(x_1 \ldots x_{m+1} \mid u, v) g(u, v \mid \alpha_1, \beta_1, \alpha_2, \beta_2) du\, dv \qquad (5)$$

$$= C_1 B(\alpha_1 + n_{01}, \beta_1 + n_{00}) B(\alpha_2 + n_{10}, \beta_2 + n_{11})$$

where

$$C_1 = \frac{P(X_1 = x_1)}{B(\alpha_1, \beta_1) B(\alpha_2, \beta_2)}.$$

Therefore, the posterior joint density of U and V, given a sample $\mathbf{X} = (x_1, \ldots, x_{m+1})$, is of the form

$$g(u, v \mid x_1, \ldots, x_{m+1}, \alpha_1, \beta_1, \alpha_2, \beta_2)$$

$$= \frac{f(x_1, \ldots, x_{m+1} \mid u, v) g(u, v \mid \alpha_1, \beta_1, \alpha_2, \beta_2)}{f(x_1, \ldots, x_{m+1} \mid \alpha_1, \beta_1, \alpha_2, \beta_2)} \qquad (6)$$

$$= C_2 u^{\alpha_1 + n_0 - 1}(1 - u)^{\beta_1 + n_{00} - 1} v^{\alpha_2 + n_{10} - 1}(1 - v)^{\beta_2 + n_{11} - 1}$$

where

$$C_2 = \frac{1}{B(\alpha_1 + n_{01}, \beta_1 + n_{00}) B(\alpha_2 + n_{10}, \beta_2 + n_{11})}.$$

162

This implies that the posteriors of U and V are conditionally independent, given a sample sequence of observations $(x_1, \ldots, x_{m+1})$. Moreover, the posterior distribution of U is Beta $(\alpha_1 + n_{01}, \beta_1 + n_{00})$, and the posterior distribution of V is Beta $(\alpha_2 + n_{10}, \beta_2 + n_{11})$, given $\mathbf{X} = (x_1, \ldots, x_{m+1})$. Hence, the posterior means and variances of U and V are:

$$E(U \mid \mathbf{X}) = \frac{\alpha_1 + n_{01}}{\alpha_1 + \beta_1 + n_{01} + n_{00}}$$

$$E(V \mid \mathbf{X}) = \frac{\alpha_2 + n_{10}}{\alpha_2 + \beta_2 + n_{10} + n_{11}} \tag{7a}$$

$$\mathrm{Var}(U \mid \mathbf{X}) = \frac{(\alpha_1 + n_{01})(\beta_1 + n_{00})}{(\alpha_1 + \beta_1 + n_{01} + n_{00})^2(\alpha_1 + \beta_1 + n_{01} + n_{00} + 1)}$$

$$\mathrm{Var}(V \mid \mathbf{X}) = \frac{(\alpha_2 + n_{10})(\beta_2 + n_{11})}{(\alpha_2 + \beta_2 + n_{10} + n_{11})^2(\alpha_2 + \beta_2 + n_{10} + n_{11} + 1)}. \tag{7b}$$

Employing an empirical Bayes approach, we want to replace in equation (7) the unknown parameters α_i and β_i by their corresponding MLE estimators $\hat{\alpha}_i$ and $\hat{\beta}_i$, which may be obtained numerically from a likehood function for r patients based on equation (5).

INCOMPLETE DATA

Maximum Likelihood Estimates

Maximum likelihood methods are commonly used to analyze Markov-chain data. The asymptotic properties of maximum likelihood estimators and likelihood ratio tests for complete data have been studied by Anderson and Goodman (1957) and Billingsley (1961), among others. Aalen and Johansen (1978) considered nonhomogeneous Markov chains with censored observations. Keiding and Gill (1990) discussed random truncation models for Markov processes. Muenz and Rubinstein (1985) discussed cases where missing values are present. This section investigates the case where there are missing values in the binary sequence of urine sample results.

Because a transition through $(k - 1)$ consecutive missing values contributes a k-step transition probability to the observed likelihood, we can calculate the k-step transition probabilities in terms of elements of the one-step transition probabilities. Therefore, it is natural to adjust the likelihood function as follows. Let

denote the two-state transition matrix considered in section 2. Then $\mathbf{P}^k = (p_{ij}^{(k)})$ denotes the matrix of k-step transition probabilities. Let $n_{ij}^{(k)}$, for $i, j = 0, 1$, denote the number of transitions from state i to state j with $(k - 1)$ consecutive missing values in between. Then, conditional on the first test result $X_1 = x_1$ the log-likelihood function for an individual can be written as

$$\log f(x_1, \ldots, x_{m+1} \mid p_{01}, p_{10}) = C + \sum_{k=1}^{m} \sum_{i=0}^{1} \sum_{j=0}^{1} n_{ij}^{(k)} \log(p_{ij}^{(k)}) \tag{8}$$

where

$$
\begin{aligned}
p_{00}^{(k)} &= (p_{01} + p_{10})^{-1}[p_{10} + p_{01}(1 - p_{01} - p_{10})^k] \\
p_{01}^{(k)} &= (p_{01} + p_{10})^{-1}[p_{01} - p_{01}(1 - p_{01} - p_{10})^k] \\
p_{10}^{(k)} &= (p_{01} + p_{10})^{-1}[p_{10} - p_{10}(1 - p_{01} - p_{10})^k] \\
p_{11}^{(k)} &= (p_{01} + p_{10})^{-1}[p_{01} + p_{10}(1 - p_{01} - p_{10})^k].
\end{aligned}
\tag{9}
$$

To simplify notation, let u, v denote the one-step transition probabilities p_{01} and $p10$, respectively. The conditional log-likelihood function can be expressed as follows:

$$
\begin{aligned}
&\log f(x_1, \ldots, x_{m+1} \mid u, v) \\
&= C - \sum_{k=1}^{m} \sum_{i=0}^{1} \sum_{j=0}^{1} n_{ij}^{(k)} \log(u + v) + \sum_{k=1}^{m} n_{00}^{(k)} \log[v + u(1 - u - v)^k] \\
&\quad + \sum_{k=1}^{m} n_{01}^{(k)} \log[u - u(1 - u - v)^k] + \sum_{k=1}^{m} n_{10}^{(k)} \log[v - v(1 - u - v)^k] \\
&\quad + \sum_{k=1}^{m} n_{11}^{(k)} \log[u + v(1 - u - v)^k]. \\
&= C + \sum_{k=1}^{m} \sum_{i=0}^{1} \sum_{j=0}^{1} n_{ij}^{(k)} \Big\{ -\log(u + v) + \\
&\qquad\qquad\qquad \log\big[u^j v^{1-j} + (-1)^{i+j} u^{1-i} v^i (1 - u - v)^k\big] \Big\}
\end{aligned}
$$

where C is a constant.

Estimates for each patient for the maintenance phase of the study are calculated. MLEs of u and v, denoted by $\hat{u}$ and $\hat{v}$, can be obtained numerically by the Newton-Raphson method. Therefore, $\hat{p}_1 = \hat{u}/(\hat{u} + \hat{v})$, estimate of the probability of using opiate of each patient, can be obtained.

For weeks 1 through 17, $\hat{P}_1$, MLE of the probability of using opiate for patients assigned with buprenorphine, methadone 20 mg, and methadone 60 mg have means (standard deviation) given by 0.4734 (0.3719), 0.6288 (0.3616) and 0.4970 (0.3157), respectively. These figures are listed in table 1. Note that patients who have only one urine sample result do not contribute any information in the study of Markov transition probabilities and, hence, are not included in the computation.

Due to the fact that there are many cases where the length of the binary sequence is very short, weighted means of P_1 are calculated in table 2. The weights used in the calculation are the numbers of informative transitions, that is, transitions that do not begin or end with a missing value, in the sample sequences. The weighted means of $\hat{P}_1$ are 0.3664 for buprenorphine, 0.6260 for methadone 20 mg, and 0.4854 for methadone 60 mg.

MLEs of U, V, and P_1 vary widely. This phenomenon, however, may be attributed to the heterogeneity of individuals involved in the study. To take into account the variability among individuals within each treatment group, an empirical Bayes approach is undertaken. By adjusting Anderson and Goodman's (1957) results for complete data, procedures for testing the homogeneity for a data set with missing values will be investigated. Censored cases can be incorporated by modifying Aalen and Johansen's (1978) method. This will be done in a subsequent paper.

TABLE 1. *Summary statistics for P1. The probability of using opiate.*

Method	Number of Cases	Mean (standard deviation)
Buprenorphine	# of cases=50	0.4734 (0.3719)
Methadone 20 mg	# of cases=53	0.6288 (0.3616)
Methadone 60 mg	# of cases=52	0.4970 (0.3751)

TABLE 2. *Summary statistics of P1. The probability of using opiate. Weighted mean. (The weights used in the calculation are the numbers of informative transitions in the sample sequences.)*

Method	Mean (standard deviation)
Buprenorphine	0.3664
Methadone 20 mg	0.6260
Methadone 60 mg	0.4854

LIMITATIONS AND DISCUSSION

1. In the section titled Incomplete Data, random variables U and V are assumed to have independent Beta priors. Similar empirical Bayes inferences can be derived if U and V are assumed to have dependent joint distributions.

2. An alternative method to incorporate missing values in a Markov model is to consider the missing value as a third state in the process. Thus, one can transform the problem into a three-state Markov model with complete data. In this case independent Dirichlet priors should be used, instead of independent Beta priors, for entries in the one-step transition probabilities matrix. It is necessary, however, to justify the randomness of the missing state.

3. For a better application of the Markov chain model for urine testings, a systematic sampling scheme with an equal number of days in between tests should be planned. This can be done by taking urine samples on Monday, Wednesday, and Friday for one week and then on Tuesday and Thursday the next week. Continuing to take samples in this 2-week alternating pattern will best use the proposed Markov model, which can handle missing observations for Sunday and Saturday in alternating weeks.

REFERENCES

Aalen, O.O., and Johansen, S. An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand J Statist* 5:141-150, 1978.

Anderson, T.W., and Goodman, L.A. Statistical inference about Markov chains. *Ann Math Stat* 28:89-110, 1957.

Basawa, I.V., and Prakasa Rao, B.L.S. *Statistical Inference for Stochastic Processes.* London: Academic Press, 1980.

Billingsley, P. *Statistical Inferences for Markov Processes.* Chicago: University of Chicago Press, 1961.

Cox, D.R.. and Miller, H.D. *The Theory of Stochastic Processes.* London: Methuen, 1965.

Keiding, N., and Gill, R.D. Random truncation models and Markov processes. *Ann Stat* 18:582-602, 1990.

Martin, J.J. *Bayesian Decision Problems and Markov Chains.* New York: Wiley, 1967.

Muenz, L.R., and Rubinstein, L.V. Markov models for covariate dependence of binary sequences. *Biometrics* 41:91-101, 1985.

ACKNOWLEDGMENTS

AUTHOR

Mei-Ling Ting Lee, Ph.D.
Statistics Department
Harvard University
One Oxford Street
Cambridge, MA 02138

Summary of Discussion: "A Markov Model for NIDA Data on Treatment of Opiate Dependence" by Mei-Ling Ting Lee*

Alan J. Gross

In her chapter, Dr. Mei-Ling Ting Lee considers a two-state Markov chain to model urine sample test results in which the outcome IS either positive or negative for the presence of a particular opiate. At times, study subjects do not appear for these tests, and as a result, missing values are present in the data. These missing values are taken into account by adjusting the likelihood function using Markov properties.

The vector of observations on a given subject is assumed to form a sequence of Markov-dependent Bernoulli trials with time-independent transition probabilities, The Markov property allows formation of the likelihood function based on a sample sequence of test results of a given subject. Maximum likelihood estimates are obtained for the one-step transition probabilities, and then, assuming stationarity, an estimate of the stationary drug abuse probability of each subject is obtained. Since these transition probabilities vary among subjects, their variability is modeled by assuming that ß prior-density functions govern their behavior. However, this aspect of the research has not been totally developed. Finally, an estimate for the stationary drug abuse probability is obtained over each of the three treatment groups.

Some generalizations that may be considered include (1) development of a three-state model in which the third state is for the missed visits by subjects (this may be preferable to the present model since missing observations currently are treated by multiplying the transition matrix to the power corresponding to the number of consecutive missing observations) and

(2) relaxation of time homogeneity and equally spaced observations currently required in the model.

Some difficulties that exist in applying this model are:

- The Markov assumption-that is, the probability an individual tests drug-free on the current test depends only on whether he or she tested drug-free on his or her immediate past test-is questionable. One testing drug-free is likely to have tested drug-free on more than just one previous test, that is, several tests back.

- Much information is likely to lie within the missing data. This aspect of the model development process is where future efforts should be placed.

- If, indeed, the transition probabilities are time homogeneous, then times between transitions should be roughly exponential. Thus, a test for exponentiality should be considered to check this assumption.

- Concern exists as to whether the data have been sampled for a sufficiently long period to assume stationarity. At this point, the assumption may be somewhat optimistic.

Dr. Mei-Ling Lee's chapter describes an interesting and potentially useful method for analyzing clinical trial data with this structure. Modifications of the procedure presented herein, once they are implemented, will lead to an improved model.

AUTHOR

Alan J. Gross, Ph.D.
Professor of Biostatistics
Department of Biostatistics, Epidemiology, and Systems Science
Medical University of South Carolina
171 Ashley Avenue
Charleston, SC 29425-2503

Open/Panel Discussion: Analysis Issues

Ram B. Jain

Panel Members: Carol K. Redmond (chair), Lloyd D. Fisher, Dean Follmann, Joel B. Greenhouse, Alan J. Gross, and A.S. Hedayat

The primary issues discussed during this discussion session were:

- Statistical approaches to analyze urine data

- Validity and importance of craving scores, physician/staff/patient global scores, and withdrawal symptoms and signs data

- Treatment of missing/censored observations

STATISTICAL APPROACHES TO ANALYZE URINE DATA

The following statistical approaches to analyze urine data were discussed:

1. Parametric and nonparametric methods to calculate summary statistics, for example, estimate of p across clinic visits

2. Model-based, nonsurvival type methods, including those based on Markovian theory

3. Single- and multiple-failure survival methods

It was mentioned that the statistical approach to analyze urine data would depend on the nature of the primary outcome variable in the study. A fair number of participants were in favor of analyzing these data by parametric model-based approaches (e.g., one proposed by Follmann et al., this volume), including those using Markovian theory because these models can provide for (1) explicit consideration of missing/censored observations, (2) inclusion of covariates representing design and population characteristics, and (3) consideration of other, for example, sociological variables, which may

help to better understand the total phenomenon of drug dependence. Dr. Hedayat was of the view that if we are interested in implications of our actions in the future, the model-based approach should be the approach of choice.

There were others (e.g., Dr. Follmann) who thought that calculation of summary statistics, such as across-clinic visits, is simple and straightforward and is sufficient to go through the Food and Drug Administration (FDA) approval process. According to Dr. Follmann, if a model must be used, it should be restricted to modeling missing/censored observations. Correlations and/or interactions between other variables, for example, correlations between day (Monday vs. Wednesday) of visit and urine sample results, are not important enough to get the drugs approved by FDA. Modeling is more suitable to describe patterns of behavior. Dr. Gross preferred some combination of summary statistics such as p and a model allowing for covariates where a model would be superimposed on, for example, p. Dr. Greenhouse was interested in a problem-driven approach. According to him, heterogeneity is a serious and most important issue in these trials, and an approach that deals with these kind of issues, for example, one that is based on Empirical Bayes' methods (Lee, this volume), would be preferable.

Dr. Jack C. Lee reminded that drug addiction is a chronic condition. He thought summary statistics and modeling that analyze response are more appropriate for analyzing data generated by studies in acute conditions. In an area such as drug dependence, application of survival-type methods should not be summarily dismissed and should be considered as in other chronic diseases (e.g., cancer) where it is more important to study survival than response. For Dr. Follmann, it was more a matter of taste. If survival-type methods are to be considered for these trials, there would be many remissions and relapses.

There was a fair amount of discussion about the pros and cons of a hypothesis-testing vs. a model-based approach for analyzing these trials. Dr. Follmann and others were in favor of formulating one or more (very) specific primary hypotheses that will help get the drug approved by FDA. Modeling is more suitable for describing the pattern of behavior. When reminded that these trials result in possible multiple endpoints, it was said that as long as hypotheses are prespecified, be it a (linear) combination of several things and/or variables, it would be in the best interest of getting a drug approved by FDA. However, if several endpoints have to be integrated, it should be described in advance how they would be integrated and the rationale for doing so. Dr. Hedayat thought that hypothesis testing generates binary (e.g., yes vs. no) results that are too restricted. He thought that scientifically it is more important to know, understand, and learn from the process/phenomenon and that this can be done

by modeling the process. It was said that FDA has to make a decision and that a physician, when treating an individual patient, must know whether the drug works. However, the information that hypothesis testing may provide to the physician may be too little. He or she not only wants to know whether or not the drug works but also how, for example, to titrate the drug for different patients (e.g., a child vs. an adult). Dr. Greenhouse noted that there is a place for Bayesian and decision theory in designing and analyzing these trials because these approaches result in meaningful measures such as the probability of clinically effective response. "These sorts of methods give us a formalism for doing sensitivity analysis to assess the robustness of our analyses," he said. Understanding and learning from the process is important and necessary for future benefits, but a decision also must be made within the confines of what is known and knowable now.

There was some discussion about what research should be done to develop better methods of analyzing these trials. It was felt that the process of missing/censoring should be modeled. There was also a strong feeling to get more and better data on those who drop out. When it was mentioned that subjects who missed three consecutive visits in the ARC 090 study were dropped, Dr. Fisher said if they had wanted to come back they should have been welcomed back into the study. This would have generated more data. Dr. Rolley E. Johnson explained the practical difficulties in putting these people back in the study, including the problem of titrating them back to their assigned doses, the ethical problem of putting them back on opiate treatment if they have gone through the worst part of the withdrawal process, and the possibility of blind being broken during the period they were out of the trial. There is also the difficulty of analyzing data on these patients if they are reentered into the trial. Should the postre entry data on these patients be integrated with the rest of the data? Or should these data be analyzed separately, and if so, what conclusions can be drawn from these data? Should the last observation be carried forward for these patients for the purpose of analyzing the main data set? Some of the patients who dropped out may not want to come back or should not be allowed to come back at all because treatment was a failure for them or because treatment cured them and they did not need the treatment any more.

A suggestion that incentives be built into the design to improve compliance (i.e., dropout rate) resulted in an involved, informative discussion. Doubts were raised about the appropriateness of introducing incentives to improve dropout rates. Incentives may affect the outcome variable itself, in which case it may not be known whether one is evaluating the effect of incentives or the treatment. Or there might be an incentive therapy interaction. This will depend on what the incentives are tied to. For example, if the incentives, particularly the financial incentives, are tied to producing a negative urine,

then it will result in informative missed visits (i.e., a clinic visit might be missed because a positive urine may be detected), or financial incentive, rather than treatment, might be the factor to promote absence from drug use. Dr. Fisher recognized that there might be an interaction between therapy and incentives and that incentive also may affect compliance, but, "nevertheless, a beneficial differential drug effect on the top of money would probably mean some sort of beneficial drug effect." A concern that different treatment groups may somehow be imbalanced in terms of incentives received by them could be rectified if patients can be stratified "on some covariate which would reflect their propensity or likelihood to take a financial incentive," said Dr. Follmann. Dr. Johnson thought that such incentives/contingencies have the potential to dilute the treatment differences, resulting in prohibitive sample size requirements. Dr. Peter A. Lachenbruch was concerned about using financial incentives because this money might end up being used for buying illegal drugs. Dr. Hedayat insisted on putting more emphasis on characterizing patient populations to blend the concepts of sampling and design and to come up with robust designs that minimize natural deficiencies, rather than, for example, create contingencies to strive for "noise-free" (e.g., dropouts) experiments. Dr. Gross mentioned that randomized response models that have a well-developed theory might be suitable for implementing Dr. Hedayat's recommendations. However, Dr. Greenhouse reminded that randomized response models would answer a very limited question in the context of clinical trials in the drug abuse area.

VALIDITY AND IMPORTANCE OF CRAVING SCORES, PHYSICIAN/STAFF/ PATIENT GLOBAL SCORES, AND WITHDRAWAL SYMPTOMS AND SIGNS DATA

Dr. Greenhouse suggested that some of these variables may be independent or explanatory variables, unrelated to the pharmacological effect of the treatment; as such, they would be time-varying covariates. If so, this presents challenging statistical analysis problems.

Dr. Fisher expressed that physician/staff/patient global scores are very crude measures and often do not work. Dr. Michael Murphy replied that,

> the approval of Alzheimer's drugs has now basically stopped
> in its tracks because no one has been able to show global
> improvement rating. . . . It depends on how one does the
> global . . . .

They range from bad to worse, and there is a big controversy right now in the field as to how one might get at that very useful information.

As Dr. Charles Gorodetzky put it, these are difficult-to-measure behavioral endpoint variables. However, these are the variables that answer questions such as "what have you done to help the patient?" These are difficult issues, but they must be dealt with.

A need was felt about some sort of global summary measure that can describe the status of patients in a global sense, for example, how he or she is doing after the trial is over. Dr. Nancy L. Geller pointed out that an optimal linear combination of multiple endpoints can be obtained; that is, a test of missingness can be combined with a test for efficacy. Weights assigned to different endpoints can be determined by experts in the field (e.g., cardiologists in cardiovascular trials). This methodology is very powerful and can detect relatively smaller differences. However, interpretation of these weighted summary measures may not always be easy.

Dr. Donald R. Jasinski thought craving scores were not a valid measure of any form of chemotherapeutic effect and probably had nothing to do with the whole process of drug dependence. In addition, there was no agreed-on meaning of craving. "It is probably related to the environment and to learned behavior . . . ," Dr. Jasinski said.

Dr. Jasinski was also concerned about moving away from dichotomous measures to continuous or ordinal scale measures to do some sort of parametric analysis. He thought all these measures, such as global scores and symptoms and signs measures, could as well be measured on a binary scale, which probably would be as meaningful as measuring them on a continuous or ordinal scale.

Drug dependence is a chronic relapsing disorder. As in other trials of chronic diseases (e.g., oncology trials), a partial or complete remission (as determined by urine data) is good enough to show that the drug is efficacious. The key issue in these trials is to establish pharmacological efficacy. It is very difficult to show differences in behavioral outcome measures.

TREATMENT OF MISSING/CENSORED OBSERVATIONS

I asked the panel which of the several methods of treating missing/censored observations presented in the meeting is more acceptable than others. The methods presented for consideration were estimate, substitute (e.g., by 1), model, and ignore. Dr. Lee observed estimate, substitute, and ignore to be different forms of modeling, and as such, modeling and missing at random are the only alternatives. Dr. Lachenbruch wanted missingness to be considered as a response by itself.

Dr. Fisher recommended doing a sensitivity analysis where different values for missing/censored observations on a range from zero (a negative sample) to one (a positive sample) are assumed and different values for different treatments are assigned. Thus, for each pair of treatment arms, a probability matrix can be constructed. Each cell of this probability matrix would provide the significance level, or p value, available from test of treatment differences when specific pairs of values attached to that cell are assigned to missing/censored observations for the two treatment groups. From this matrix, a set of pairs of values (assigned to missing/censored observations) for which one treatment can be inferred to be better than others can be determined. A professional judgment then can be made as to whether, under reasonable assumptions for missing/censored data, one treatment is better. This approach provides for scenarios under which conclusions could be different.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical   Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Open/Panel Discussion: General Issues

Ram B. Jain

Panel Members: Peter A. Lachenbruch (Cochair), Jack C. Lee (Cochair), Joseph Collins, Lloyd D. Fisher, Sudhir G. Gupta, Nicholas P. Jewell, Michael Murphy, Vincent Shu, and Ram C. Tiwari

INTRODUCTION

The issues primarily discussed during this session were:

- Need for multiple outcome criteria to deal with various aspects of drug addiction

- High and differential dropout rates and their influence on estimation of treatment effect

- Alternate study designs: comparative dose trials and "enrichment" designs

NEED FOR MULTIPLE OUTCOME CRITERIA TO DEAL WITH VARIOUS ASPECTS OF DRUG ADDICTION

Almost all the participants at the meeting agreed that there is a need for multiple outcome criteria in these trials. However, what should be measured should be clearly defined. Dr. Murphy would definitely incorporate urine screens as a measure in his studies, particularly because it would be easy to standardize this measure across the centers. However, this would not be the key, pivotal outcome in his studies because this measure is of "questionable clinical relevance." He would like to study the impact of urine screens "on the sites we select, the patients we enroll, the numbers of patients we enroll, and how often we can retain them." If the patients become free of opiates but substitute opiates by using other illicit drugs like Valium, cocaine, amphetamine, or marijuana, he would consider this a failure of (opiate) treatment. This substitution of other drugs for opiates should be factored into data analysis.

176

Dr. Gupta warned against having too many outcome measures because correlations between these measures must be considered in the analysis. Too many measures make the problem multidimensional and too complicated. Obtaining a clear, solid conclusion from the trial is important. However, because the effect of (opiate) treatment drugs lasts for only a limited time–24 hours or so-he was in favor of analyzing data collected on each day separately. According to Dr. Jewell, the primary interest should be evaluating whether the patient has improved since the treatment began, and as such, treating multiple observations from a patient as anything more than a single data point should not be done.

There was no consensus at the meeting about how to handle the data on multiple outcome variables. Dr. Murphy was in favor of integrating data from urine screens with other measures into an overall index of disease severity. He liked the idea of collapsing across measures. Dr. Collins wanted each measure to be analyzed separately because, in general, investigators may be interested in knowing exactly on which measures what the direction and magnitude of treatment differences are, if any exist. A drug may reduce the opiate abuse but may have a problem in retaining patients on the treatment. On the other hand, a drug may not be so effective in reducing the frequency of drug abuse but may be very effective retaining patients on the treatment. The investigator should be the one to make a final decision about the usefulness of the proposed treatment.

However, as Dr. Fisher pointed out, if there are too many outcome measures, it is more than likely that the direction and magnitude of treatment differences *on* these variables will be different at different levels. This could be a potential source of confusion and indecision for the investigator. It would only be helpful if a composite measure is obtained from different outcome measures. The analysis of this combined measure should be helpful in deciding if the drug has an overall effect. Subsequently, individual measures can be analyzed and evaluated. I suggested that an investigator be asked at the beginning of a trial about the performance required on each outcome measure for the proposed drug to be minimally useful and/or successful. For example, for the proposed drug to be useful and/or successful, patients must participate in the study for at least 4 weeks, should have at least 75 percent negative urines, and should have a craving score of no more than 50 on at least 75 percent of their clinic visits. The performances on different measures for each patient can then be combined to obtain some sort of combined score without statistical manipulations. These combined scores can then be subjected to formal analysis.

Dr. Fudala did not think a single composite score would be clinically meaningful. Dr. Shu expressed the need for a more focused study and the need for defining a primary outcome measure, for example, the reduction in incidence of drug abuse. Other secondary measures, such as rating scales to measure patients' physical or psychological dependence on drugs, can be developed. Craving scores, a nonthreatening measure, can also be used. Correlations between craving scores and incidence of drug abuse can be studied. The bias in obtaining some of the soft measures can be reduced and/or eliminated if blinded evaluating physicians, separated from treating physicians, take these measures.

As Dr. Geller put it, there are likely to be situations where combining several measures (test statistics) would be appropriate, and in some cases, it would be more appropriate to have one primary and several secondary measures (statistics).

HIGH AND DIFFERENTIAL DROPOUT RATES AND THEIR INFLUENCE ON ESTIMATION OF TREATMENT EFFECT

Almost all participants were concerned about the high dropout rates in these trials. Dr. Jewell thought, with that much missing/dropout data "one can say anything about treatment effect." In "a trial of this kind . . . where 80 percent of the patients on placebo were not available with regard to any kind of outcome or restricted . . . outcome, there is just no way one would be able to make sense of what that meant with regard to an active drug . . . ." He advised that pilot studies that concentrate on the ability to collect outcome data of specified kinds be conducted so that, for instance, specific patient populations available for specific treatments and followup can be identified.

There was an agreement that high and differential dropout rates for different reasons in different treatment groups can compromise estimation of treatment effects. For example, the drugs may "cure" certain patients, and as such, they do not need treatment anymore (and they drop out). In other cases, the drug may be a failure, and as such, the patients do not come back. Some patients may drop out because of toxicity even if the drug was working for them. Other times the pattern of dropouts may be related to covariates such as age, sex, and marital status.

In the opinion of Dr. Gupta, dropout rates may contain information about the treatment effect, and as such, dropout rates should be analyzed as a separate outcome variable in addition to other variables related to the improvement in the condition of patients. Dr. Jack C. Lee also said, ". . . if the missingness has any valuable information for the relevance of the outcome variable, which I think in

this case is, then make it a part of the analysis . . . ." According to Dr. Murphy, only legitimate discontinuation from the study is due to toxicity. Dropouts have "only a fraction to do with the therapeutic intervention in place. . . . it is an issue of staff investment." Clinical staff should prevent that from happening. Dr. Johnson commented that an administrative dropout after three consecutive missed visits may be one of the reasons for a large dropout rate in the ARC 090 trial, but ". . . when you don't put a restriction on with medications that have physical dependence-producing properties you bring in a lot of complications that I don't know how you deal with . . . ." Dr. Follmann suggested that efforts could have been made to collect urine samples from the patients even after they were out of the trial. Dr. Johnson retorted that they (patients) could have been paid for coming back to provide urine samples but that a provision like that would constitute payment for dropping out of the trial, which brings in an ethical problem because these payments could be used to buy drugs of abuse. Dr. Johnson remarked, ". . . how do you get people back when they did not come to begin with?" Missed visits is an inherent phenomenon in these patient populations.

It was suggested that sensitivity analysis recommended by Dr. Fisher and the approaches proposed by Drs. Weng (this volume) and Follmann and colleagues (this volume) are promising tools to adjust for missing/censored data.

ALTERNATE STUDY DESIGNS

Comparative Dose Trials

Most of the participants recognized the usefulness of comparative dose trials. However, two issues must be dealt with in these kinds of designs. First, the selection of doses should be done carefully. As Dr. Murphy said, "Early in the development especially, it is very difficult to judge where one is on the dose effect curve. One could be asymptotically too high or too low." Dr. Shu suggested the physicians should be allowed to titrate in the first studies to select the doses they feel comfortable with, and these selected doses can be used in dose comparison trials. The second difficulty in doing these kinds of trials lies in that, in the neuropsychiatric area, the effect sizes are very small. Relatively large sample sizes may be required to detect differences between these small effects for different treatment groups.

However, Dr. Fisher was, in general, not in favor of doing forced dose escalation studies.

179

"ENRICHMENT" DESIGNS

It would be of interest to physicians to minimize the damage to those who cannot benefit from a proposed new therapy. Those who can benefit probably can be identified from those who cannot in a double blind trial where everybody receives medication. Then, in a random double blind second (e.g., dose comparison) trial, or probably the second phase of the same trial, only the responders can be studied. At the request of Dr. Donald F. Klein, professor of psychiatry at the College of Physicians and Surgeons of Columbia University, I asked the participants to comment on the usefulness of such designs in the drug abuse area.

Dr. Jewell thought the main idea behind these kinds of designs is to identify subgroups based on some sort of covariate information that amounts to evaluation of interaction effects. If so, these trials could result in large sample size requirements. Dr. Fisher said that these trials are initiated when physicians intuitively feel that some substrata of patient population are biologically and/or psychologically so different from the others that for some there is no hope, whereas others can respond. This is an efficient way to show activity. As Dr. Jack C. Lee said, these designs can be useful in determining the profiles of the drug addicts who respond. This information can then be used to determine eligibility for future studies. These designs can be used to demonstrate heterogeneity between responders and nonresponders, but this heterogeneity, if present, has to be rooted, in Dr. Murphy's opinion, in some sort of biological substrate. It should flow from theory into practice.

Several participants expressed reservations and struck a note of caution about the use of "enrichment" designs. The analysis of these designs is difficult because of, among other things, possible carryover effects. The interpretation of results obtained by analyzing these trials presents serious challenges. Dr. Vocci presented two examples where one can incorrectly conclude that an alternate medication is ineffective. "If you give an agonist and find out who responds to the agonist, and then you are testing an agonist in an enrichment design vs. a partial agonist, what you may be doing is undermedicating the patients with partial agonist and making ii look worse . . . . A study was done in which anxiety patients were tested for benzodiazepine receptor, benzodiazepine sensitivity, if they were actually helped by benzodiazepine, and then they were randomized to the same benzodiazepine or a nonbenzodiazepine, and what happened was the patients who were randomized to the nonbenzodiazepine had this mild withdrawal, which was not helped by the new agent, and many of them dropped out as a result. . . . This withdrawal syndrome actually exacerbated their anxiety." Dr. Greenhouse presented a similar example from a trial done in the psychiatric area.

Consequently, although enrichment designs do have a place in the drug development process (particularly when a new chemical entity is being studied), where a post hoc analysis of the data will be meaningful in identifying easily discernible (by the average physician), clinically relevant baseline variables that separate responders from nonresponders, prospectively, that "good" group should be studied in a traditional clinical trial.

AUTHOR

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Technical Review Committee Meeting: Statistical Issues in Clinical Trials for Treatment of Opiate Dependence

Medications Development Division
National Institute on Drug Abuse
December 2-3, 1991

List of Participants

Jack D. Blaine, M.D.
Chief
Treatment Research Branch
Division of Clinical Research
National Institute on Drug Abuse
Parklawn Building, Room 10A-30
5600 Fishers Lane
Rockville, MD 20857

Robert J. Chiarello, M.D.
Medical Officer
Treatment Research Branch
Division of Preclinical Research
National Institute on Drug Abuse
Parklawn Building, Room 10A-30
5600 Fishers Lane
Rockville, MD 20857

Joseph Collins, Sc.D.
Chief
Cooperative Studies Program
  Coordinating Center, 151E
Department of Veterans Affairs Medical
  Center
Perry Point, MD 21902

Edward J. Cone, Ph.D.
Chief
Laboratory of Chemistry and Drug
  Metabolism
Addiction Research Center
National Institute on Drug Abuse
P.O. Box 5180
Baltimore, MD 21224

Sandra L. Dickerson, B.S.
Medical Technologist
Addiction Research Center
National Institute on Drug Abuse
P.O. Box 5180
Baltimore, MD 121224

Lloyd D. Fisher, Ph.D.
Professor
Department of Statistics
SC-32
University of Washington
19220 64th Place, N.E.
Seattle, WA 98155

Dean Follmann, Ph.D.
Mathematical Statistician
Biostatistics Research Branch
National Heart, Lung, and Blood
  Institute
Federal Building, Room 2A11
Bethesda, MD 20892

Paul J. Fudala, Ph.D.
Assistant Professor
Department of Psychiatry
University of Pennsylvania School of
  Medicine and the Department of
  Veterans Affairs Medical Center
Building 15
University and Woodland Avenues
Philadelphia, PA 19104

Nancy L. Geller, Ph.D.
Chief
Biostatistics Research Branch
National Heart, Lung, and Blood
  Institute
Federal Building, Room 2All
Bethesda, MD 20892

Albert J. Getson, Ph.D.
Associate Director
Merck Clinical Biostatistical Research
  Labs
BL 3-2
West Point, PA 19486

Harold Gordon, Ph.D.
Project Officer
Division of Epidemiology and
  Prevention Research
National Institute on Drug Abuse
5600 Fishers Lane
615 Rockwall II
Rockville, MD 20852

David A. Gorelick, M.D., Ph.D.
Chief
Treatment Research Branch
Addiction Research Center
National Institute on Drug Abuse
P.O. Box 5180
4940 Eastern Avenue
Baltimore. MD 21224

Charles W. Gorodetzky, M.D., Ph.D.
Executive Director
CNS Drug Development
CIBA-Geigy Corporation
556 Morris Avenue
Summit, NJ 07901

Joel B. Greenhouse, Ph.D.
Associate Professor
Department of Statistics
Carnegie-Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

Alan J. Gross, Ph.D.
Professor
Department of Biostatistics,
  Epidemiology, and Systems Science
Medical University of South Carolina
  College of Graduate Studies
  171 Ashley Avenue
Charleston, SC 29425-2503

Charles V. Grudzinskas, Ph.D.
Director
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Sudhir C. Gupta, Ph.D.
Associate Professor
Statistics Department
Northern Illinois University
DeKalb. IL 60115-2888

AS. Hedayat, Ph.D.
Professor
Department of Mathematics, Statistics,
  and Computer Sciences, MC249
University of Illinois
P.O. Box 438
Chicago, IL 60680

John Hyde, M.D., Ph.D.
Medical Officer
Pilot Drug Evaluation Division, HFD-007
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857

Ram B. Jain, Ph.D.
Mathematical Statistician
Biometrics Branch
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

Donald R. Jasinski, M.D.
Chief
Center for Chemical Dependence
Francis Scott Key Medical Center
4940 Eastern Avenue A-l-East
Baltimore, MD 21224

Nicholas P. Jewell, Ph.D.
Professor of Biostatistics
University of California at Berkeley
  School of Public Health
Berkeley, CA 94720

Rolley E. Johnson, Pharm.D.
Assistant Professor
Department of Psychiatry and
  Behavioral Sciences
Johns Hopkins School of Medicine
Building G, Room 2725
85510 Nathan Shock Drive
Baltimore, MD 21224

William F. Krol, Ph.D.
Chief
Statistical Programming Section, 151E
Department of Veterans Affairs Medical
  Center
Perry Point, MD 21902

Peter A. Lachenbruch, Ph.D.
Professor of Biostatistics
University of California, Los Angeles
  School of Public Health
Los Angeles, CA 90024-1772

Young Jack C. Lee, Ph.D.
Chief
Biometry and Mathematical Statistics
  Branch
National Institute of Child Health and \
  Human Development
National Institutes of Health
Executive Plaza North, Room 630
Bethesda, MD 20892

Mei-Ling Ting Lee, Ph.D.
Assistant Professor
Mathematics Department
Boston University
111 Cummington Street
Boston, MA 02215

Shou-Hua Li, Ph.D.
Statistician
Biometrics Unit, EODPP
Westwood Building, Room 533
National Institute of Dental Research
Bethesda, MD 20892

Michael Murphy, M.D.
Medical Director
Neurosciences—SBU
Hoechst Roussel Pharmaceutical, Inc.
Route 202-206 North
Sommerville, NJ 08876-1258

Taesung Park, Ph.D.
Mathematical Statistician
National Institute of Child Health and
  Human Development
National Institutes of Health
Executive Plaza North, Room 630
Bethesda. MD 20892

Carol K. Redmond, Ph.D.
Professor
Department of Biostatistics
University of Pittsburgh
318 Parran Hall
Pittsburgh, PA 15261

Saul Rosenberg, Ph.D.
Mathematical Statistician
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville. MD 20857

Vincent Shu, Ph.D.
Section Head
Clinical Statistics and System
  Development
D436 APGC
Abbott Laboratories
One North Abbott Park Road
Abbott Park, IL 60064

Richard Stein, Ph.D.
Mathematical Statistician
Food and Drug Administration
HFD-007, Room 9B45
5600 Fishers Lane
Rockville, MD 20857

Ram C. Tiwari, Ph.D.
Associate Professor
Department of Mathematics
University of North Carolina at Charlotte
Charlotte, NC 28223

Frank J. Vocci, Ph.D.
Deputy Director
Medications Development Division
National Institute on Drug Abuse
Parklawn Building, Room 11A-55
5600 Fishers Lane
Rockville, MD 20857

L.J. Wei, Ph.D.
Professor
Department of Biostatistics
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

T.S. Weng, Ph.D.
Mathematical Statistician
Division of Biometric Sciences
Office of Science and Technology
Center for Devices and Radiological
  Health
Food and Drug Administration
Suite 405
1801 Rockville Pike
Rockville, MD 20852

Curtis Wright, M.D.
Medical Officer
Pilot Drug Evaluation Division
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20852

Margaret Wu, Ph.D.
Mathematical Statistican
Biostatistics Research Branch
National Heart, Lung, and Blood
  Institute
National Institutes of Health
Federal Building, Room 2A11
Bethesda, MD 20892

National Institute on Drug Abuse Research

MONOGRAPH SERIES

While limited supplies last, single copies of the monographs may be obtained free of charge from the Office for Substance Abuse Prevention's National Clearinghouse for Alcohol and Drug Information (ONCADI). Please contact ONCADI also for information about availability of coming issues and other publications of the National Institute on Drug Abuse relevant to drug abuse research.

Additional copies may be purchased from the U.S. Government Printing Office (GPO) and/or the National Technical Information Service (NTIS) as indicated. NTIS prices are for paper copy; add $3 handling charge for each order. Microfiche copies are also available from NTIS. Prices from either source are subject to change.

Addresses are:

ONCADI
National Clearinghouse for Alcohol and Drug Information
P.O. Box 2345
Rockville, MD 20852
(301) 468-2600
(800) 729-6686

GPO
Superintendent of Documents
U.S. Government Printing Office
Washington, DC 20402
(202) 275-2981

NTIS
National Technical Information Service
U.S. Department of Commerce
Springfield, VA 22161
(703) 487-4650

For information on availability of NIDA Research Monographs 1 through 70 (1975-1986) and others not listed, write to NIDA, Community and Professional Education Branch, Room 10A-54, 5600 Fishers lane, Rockville, MD 20857.

71 OPIATE RECEPTOR SUBTYPES AND BRAIN FUNCTION. Roger M. Brown, Ph.D.; Doris H. Clouet, Ph.D.; and David P. Friedman, Ph.D., eds.
GPO out of stock   NTIS PB #89-151955/AS $31

72 RELAPSE AND RECOVERY IN DRUG ABUSE. Frank M. Tims, Ph.D., and Carl G. Leukefeld, D.S.W., eds.
GPO Stock #017-024-01302-1 $6   NTIS PB #89-151963/AS $31

73 URINE TESTING FOR DRUGS OF ABUSE. Richard L. Hawks, Ph.D., and C. Nora Chiang, Ph.D., eds.
GPO Stock #017-024-01313-7 $3.75   NTIS PB #89-151971/AS $23

74 NEUROBIOLOGY OF BEHAVIORAL CONTROL IN DRUG ABUSE. Stephen I. Szara, M.D., DSc., ed.
GPO Stock #017-024-01314-5 $3.75   NTIS PB #89-151989/AS $23

75 PROGRESS IN OPIOID RESEARCH. PROCEEDINGS OF THE 1986 INTERNATIONAL NARCOTICS RESEARCH CONFERENCE. John W. Holaday, Ph.D.; Ping-Yee Law, Ph.D.; and Albert Herz, M.D., eds.
GPO out of stock   ONCADI out of stock
                                          Not available from NTIS

76 PROBLEMS OF DRUG DEPENDENCE, 1986: PROCEEDINGS OF THE 48TH ANNUAL SCIENTIFIC MEETING, THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D., ed.
GPO out of stock   ONCADI out of stock
                                          NTIS PB #88-208111/AS $53

77 ADOLESCENT DRUG ABUSE: ANALYSES OF TREATMENT RESEARCH. Elizabeth R. Rahdert, Ph.D., and John Grabowski, Ph.D., eds.
GPO Stock #017-024-01348-0 $4   ONCADI out of stock
                                          NTIS PB #89-125488/AS $23

78 THE ROLE OF NEUROPLASTICITY IN THE RESPONSE TO DRUGS. David P. Friedman, Ph.D., and Doris H. Clouet, Ph.D., eds.
GPO out of stock   NTIS PB #88-245683/AS $31

79 STRUCTURE-ACTIVITY RELATIONSHIPS OF THE CANNABINOIDS. Rao S. Rapaka, Ph.D., and Alexandros Makriyannis, Ph.D., eds.
GPO out of stock   NTIS PB #89-109201/AS $31

80 NEEDLE SHARING AMONG INTRAVENOUS DRUG ABUSERS: NATIONAL AND INTERNATIONAL PERSPECTIVES. Robert J. Battjes, D.S.W., and Roy W. Pickens, Ph.D., eds.
GPO out of stock   NTIS PB #88-236138/AS $31

81 PROBLEMS OF DRUG DEPENDENCE, 1987: PROCEEDINGS OF THE 49TH ANNUAL SCIENTIFIC MEETING, THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D., ed.
GPO Stock #017-024-01354-4 $17   NTIS PB #89-109227/AS
                                         Contact NTIS for price

82 OPIOIDS IN THE HIPPOCAMPUS. Jacqueline F. McGinty, Ph.D., and David P. Friedman, Ph.D., eds.
GPO out of stock   NTIS PB #88-245691/AS $23

83 HEALTH HAZARDS OF NITRITE INHALANTS. Harry W. Haverkos, M.D., and John A. Dougherty, Ph.D., eds.
GPO out of stock   NTIS PB #89-125496/AS $23

84 LEARNING FACTORS IN SUBSTANCE ABUSE. Barbara A. Ray, Ph.D., ed.
GPO Stock #017-024-01353-6 $6   NTIS PB #89-125504/AS $31

85 EPIDEMIOLOGY OF INHALANT ABUSE: AN UPDATE. Raquel A. Crider, Ph.D., and Beatrice A. Rouse, Ph.D., eds.
GPO Stock #017-024-01360-9 $5.50   NTIS PB #89-123178/AS $31

86 COMPULSORY TREATMENT OF DRUG ABUSE: RESEARCH AND CLINICAL PRACTICE. Carl G. Leukefeld, D.S.W., and Frank M. Tims, Ph.D., eds.
GPO Stock #017-024-01352-8 $7.50   NTIS PB #89-151997/AS $31

87 OPIOID PEPTIDES: AN UPDATE. Rao S. Rapaka, Ph.D., and Bhola N. Dhawan, M.D., eds.
GPO Stock #017-024-01366-8 $7   NTIS PB #89-158430/AS $45

88 MECHANISMS OF COCAINE ABUSE AND TOXICITY. Doris H. Clouet, Ph.D.; Khursheed Asghar, Ph.D.; and Roger M. Brown, Ph.D., eds.
GPO Stock #017-024-01359-5 $11   NTIS PB #89-125512/AS $39

89 BIOLOGICAL VULNERABILTY TO DRUG ABUSE. Roy W. Pickens, Ph.D., and Date S. Svikis, B.A., eds.
GPO Stock #017-022-01054-2 $5   NTIS PB #89-125520/AS $23

90 PROBLEMS OF DRUG DEPENDENCE 1988: PROCEEDINGS OF THE 50TH ANNUAL SCIENTIFIC MEETING, THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D., ed.
GPO Stock #017-024-01362-5 $17

91 DRUGS IN THE WORKPLACE: RESEARCH AND EVALUATION DATA. Steven W. Gust, Ph.D., and J. Michael Walsh, Ph.D., eds.
GPO Stock #017-024-01384-6 $10   NTIS PB #90-147257/AS $39

92 TESTING FOR ABUSE LIABILITY OF DRUGS IN HUMANS. Marian W. Fischman, Ph.D., and Nancy K. Mello, Ph.D., eds.
GPO Stock #017-024-01379-0 $12   NTIS PB #90-148933/AS $45

93 AIDS AND INTRAVENOUS DRUG USE: FUTURE DIRECTIONS FOR COMMUNITY-BASED PREVENTION RESEARCH. C.G. Leukefeld, D.S.W.; R.J. Battjes, D.S.W.; and Z. Amsel, DSc., eds.
GPO Stock #017-024-01388-9 $10   NTIS PB #90-148941/AS $39

94 PHARMACOLOGY AND TOXICOLOGY OF AMPHETAMINE AND RELATED DESIGNER DRUGS. Khursheed Asghar, Ph.D., and Errol De Souza, Ph.D., eds.
GPO Stock #017-024-01386-2 $11   NTIS PB #90-148958/AS $39

95 PROBLEMS OF DRUG DEPENDENCE 1989: PROCEEDINGS OF THE 51ST ANNUAL SCIENTIFIC MEETING, THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D., ed.
GPO Stock #017-024-01399-4 $21   NTIS PB #90-237660/AS $67

96 DRUGS OF ABUSE: CHEMISTRY, PHARMACOLOGY, IMMUNOLOGY, AND AIDS. Phuong Thi Kim Pham, Ph.D., and Kenner Rice, Ph.D., eds.
GPO Stock #017-024-01403-6 $8   NTIS PB #90-237678/AS $31

97 NEUROBIOLOGY OF DRUG ABUSE: LEARNING AND MEMORY. Lynda Erinoff, Ph.D., ed.
GPO Stock #017-024-01404-4 $8   NTIS PB #90-237686/AS $31

98 THE COLLECTION AND INTERPRETATION OF DATA FROM HIDDEN POPULATIONS. Elizabeth Y. Lambert, MS., ed.
GPO Stock #017-024-01407-9 $4.75   NTIS PB #90-237694/AS $23

99 RESEARCH FINDINGS ON SMOKING OF ABUSED SUBSTANCES. C. Nora Chiang, Ph.D., and Richard L. Hawks, Ph.D., eds.
GPO Stock #017-024-01412-5 $5   NTIS PB#91-141119 $23

100 DRUGS IN THE WORKPLACE: RESEARCH AND EVALUATION DATA. VOL. II. Steven W. Gust, Ph.D.; J. Michael Walsh, Ph.D.; Linda B. Thomas, B.S.; and Dennis J. Crouch, M.B.A., eds.
GPO Stock #017-024-01458-3 $8

101 RESIDUAL EFFECTS OF ABUSED DRUGS ON BEHAVIOR. John W. Spencer, Ph.D., and John J. Boren, Ph.D., eds.
GPO Stock #017-024-01426-7 $6   NTIS PB #91-172858/AS $31

102 ANABOLIC STEROID ABUSE. Geraline C. Lin, Ph.D., and Lynda Erinoff, Ph.D., eds.
GPO Stock #017-024-01425-7 $8   NTIS PB #91-172866/AS $31

103 DRUGS AND VIOLENCE: CAUSES, CORRELATES, AND CONSEQUENCES. Mario De La Rosa, Ph.D.; Elizabeth Y. Lambert, M.S.; and Bernard Gropper, Ph.D., eds.
GPO Stock #017-024-01427-3 $9   NTIS PB #91-172841/AS $31

104 PSYCHOTHERAPY AND COUNSELING IN THE TREATMENT OF DRUG ABUSE. Lisa Simon Onken, Ph.D., and Jack D. Blaine, M.D., eds.
GPO Stock #017-024-01429-0 $4   NTIS PB #91-172874/AS $23

105 PROBLEMS OF DRUG DEPENDENCE, 1990: PROCEEDINGS OF THE 52ND ANNUAL SCIENTIFIC MEETING, THE COMMITEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D.. ed.
GPO Stock #017-024-01435-4 $22

106 IMPROVING DRUG ABUSE TREATMENT. Roy W. Pickens, Ph.D.; Carl G. Leukefeld, D.S.W.; and Charles R. Schuster, Ph.D., eds.
GPO Stock #017-024-01439-7 $12   NTIS PB #92-105873
                                         Paperback $50 Microfiche $19

107 DRUG ABUSE PREVENTION INTERVENTION RESEARCH: METHODOLOGICAL ISSUES. Carl G. Leukefeld, D.S W., and William J. Bukoski, Ph.D., eds.
GPO Stock #017-024-01446-0 $9   NTIS PB #92-160985
                                       Paperback $35 Microfiche $17

108 CARDIOVASCULAR TOXICITY OF COCAINE: UNDERLYING MECHANISMS. Pushpa V. Thadani, Ph.D., ed.
GPO Stock #017-024-01446-0 $7   NTIS PB #92-106608
                                       Paperback $35 Microfiche $17

109 LONGITUDINAL STUDIES OF HIV INFECTION IN INTRAVENOUS DRUG USERS: METHODOLOGICAL ISSUES IN NATURAL HISTORY RESEARCH. Peter Hartsock, Dr.P.H., and Sander G. Genser, M.D., M.P.H., eds.
GPO Stock #017-024-01445-1 $4.50   NTIS PB #92-106616
                                          Paperback $26 Microfiche $12.50

110 THE EPIDEMIOLOGY OF COCAINE USE AND ABUSE. Susan Schober, Ph.D., and Charles Schade, M.D., M.P.H., eds.
GPO Stock #017-024-01456-7 $11   NTIS PB #92-14624-0
                                        Paperback $43 Microfiche $17

111 MOLECULAR APPROACHES TO DRUG ABUSE RESEARCH VOLUME I: RECEPTOR CLONING, NEUROTRANSMITTER EXPRESSION, AND MOLECULAR GENETICS. Theresa N.H. Lee, Ph.D., ed.
Not for sale at GPO   NTIS PB #92-135743
                                        Paperback $35 Microfiche $17

112 EMERGING TECHNOLOGIES AND NEW DIRECTIONS IN DRUG ABUSE RESEARCH. Rao S. Rapaka, Ph.D.; Alexandros Makriyannis, Ph.D.; and Michael J. Kuhar, Ph.D., eds.
GPO Stock #017-024-01455-9 $11

113 ECONOMIC COSTS, COST-EFFECTIVENESS, FINANCING, AND COMMUNITY-BASED DRUG TREATMENT. William S. Cartwright, Ph.D., and James M. Kaple, Ph.D., eds.
Not for sale at GPO   NTIS PB #92-155795
                                        Paperback $35 Microfiche $17

114 METHODOLOGICAL ISSUES IN CONTROLLED STUDIES ON EFFECTS OF PRENATAL EXPOSURE TO DRUG ABUSE. M. Marlyne Kilbey, Ph.D., and Khursheed Asghar, Ph.D., eds.
GPO Stock #017-024-01459-1 $12   NTIS PB #92-146216
                                        Paperback $43 Microfiche $17

115 METHAMPHETAMINE ABUSE: EPIDEMIOLOGIC ISSUES AND IMPLICATIONS. Marissa A. Miller, D.V.M., M.P.H., and Nicholas J. Kozel, MS., eds.
GPO Stock #017-024-01460-5 $4

116 DRUG DISCRIMINATION: APPLICATIONS TO DRUG ABUSE RESEARCH. Richard A. Glennon, Ph.D.; Torbjörn U.C. Järbe, Ph.D.; and Jerry Frankenheim, Ph.D., eds.
GPO Stock #017-024-01470-2 $13

117 METHODOLOGICAL ISSUES IN EPIDEMIOLOGICAL, PREVENTION, AND TREATMENT RESEARCH ON DRUG-EXPOSED WOMEN AND THEIR CHILDREN. M. Marlyne Kilbey, Ph.D., and Khursheed Asghar, Ph.D., eds.
GPO Stock #017-024-01472-9 $12

118 DRUG ABUSE TREATMENT IN PRISONS AND JAILS. Carl G. Leukefeld, D.S.W., and Frank M. Tims, Ph.D., eds.
GPO Stock #017-024-01473-7 $16

119 PROBLEMS OF DRUG DEPENDENCE 1991: 53RD ANNUAL SCIENTIFIC MEETING, THE COMMITTEE ON PROBLEMS OF DRUG DEPENDENCE, INC. Louis S. Harris, Ph.D., ed.
GPO Stock #017-024-01474-5 $22

120 BIOAVAILABILITY OF DRUGS TO THE BRAIN AND THE BLOOD-BRAIN BARRIER. Jerry Frankenheim, Ph.D., and Roger M. Brown, Ph.D., eds.

121 BUPRENORPHINE: AN ALTERNATIVE TREATMENT FOR OPIOID DEPENDENCE. Jack D. Blaine, Ph.D., ed.

122 RESEARCH METHODS IN WORKPLACE SETTINGS. Helen Axel, M.A., and Dennis J. Crouch, M.B.A., eds.

123 ACUTE COCAINE INTOXICATION: CURRENT METHODS OF TREATMENT. Heinz Sorer, Ph.D., ed.

124 NEUROBIOLOGICAL APPROACHES TO BRAIN-BEHAVIOR INTERACTION. Roger M. Brown, Ph.D., and Joseph Frascella, Ph.D., eds.

125 ACTIVATION OF IMMEDIATE EARLY GENES OF ABUSE. Reinhard Grzanna, Ph.D., and Roger M. Brown, Ph.D., eds.

126 MOLECULAR APPROACHES TO DRUG ABUSE RESEARCH VOLUME II: STRUCTURE, FUNCTION, AND EXPRESSION. Theresa N.H. Lee, Ph.D., ed.