

# Sentencing Illicit Drug Traffickers: How do the Courts Handle Random Sampling Issues?

ALAN JULIAN IZENMAN

*Temple University*

*E-mail address:* alan@temple.edu

*March 2003*

---

## ABSTRACT

During the last two years, several major developments in federal sentencing have taken place. It all started in June 2000 with *Apprendi v. New Jersey*, in which Justice O'Connor, in dissent, termed a "watershed in constitutional law." Prior to *Apprendi*, a judge would sentence a convicted drug trafficker to imprisonment using the preponderance standard of proof on the quantities of drugs seized. The affect of *Apprendi* is to make more juries decide the quantity of drugs by the reasonable doubt standard. Although *Apprendi* had nothing to do with illicit drugs, the implication now is that all federal sentencing protocols are under revision. Since *Apprendi* was decided, huge numbers of cases involving convicted drug traffickers have been appealed, with the result that a ruling from the Supreme Court on further clarification of this issue is expected. This article discusses the repercussions on various statistical issues involved in the determination of total drug quantity under the changing protocols.

*Keywords:* Controlled substances, federal sentencing guidelines, forensic statistics, homogeneity, multistage sampling, random sampling, standards of proof.

---

## 1. Introduction

Convicting a defendant of possessing an illicit drug such as heroin, powder cocaine, crack cocaine, marijuana with intent to distribute or supply (otherwise known as “trafficking” or “dealing”) is usually not difficult. The drugs are either found on the defendant, in the defendant’s home, or in some place frequented by the defendant. Circumstantial evidence, such as the possession of weapons, large amounts of cash, police scanners, beepers, cellular phones, drug paraphernalia, and the quantity of illicit drugs in the defendant’s possession and how it is packaged, also help the prosecution to establish the defendant’s intention to distribute drugs.

The courts in Britain and Europe treat the sentencing of convicted defendants in drug cases quite differently from that of the courts in the United States. British courts are not required to determine the total quantity of illicit drugs found on the defendant, except as a means of distinguishing between ‘possession’ (i.e., for personal use or supply) and ‘trafficking’. Intent, as evidenced by a number of different factors reflecting “the scale and nature of the dealing” (e.g., degree of involvement, the amount of trafficking, and the street value of the drugs being handled) is the primary determinant of sentence length (*R v Aramah*, 1982). Purity of the drugs can indicate the defendant’s closeness to the source of supply of the drugs, and the courts have found it essential to determine the level of purity in drug smuggling cases or where large quantities (at least 500 gr) of cocaine, heroin, or amphetamines are seized.

This is not to say, however, that a larger drug quantity found on a defendant in a British court may not count more heavily than a smaller amount in terms of the actual sentence imposed. In fact, recent decisions in the British courts have generally recommended that weight of the drugs (determined at 100% purity) be used as a better way of establishing the relative significance of a drug seizure than its street value; see, for example, *R v Aranguren* (1994), *R v Warren and Beeley* (1995), and *R v Morris* (2001). The argument against street price as a sentencing criterion in these cases (which involved heroin, cocaine, and Ecstasy tablets) is quite similar: if such drugs became more widely available, the street price would drop, leading to a lowered sentencing level for possession of larger drug quantities, a result clearly contrary to public policy. Nevertheless, weight of the drugs (or number of tablets) is still only one of several relevant factors used to determine the appropriate sentence.

Although the total amount of narcotic drugs and psychotropic substances is a very important factor in helping European courts comprehend the scale of drug dealing, it is typically

only one of many relevant factors used in sentencing decisions. In Austria, for example, specific limits (in grams) are set for each type of illicit drug (e.g., cocaine, 15 grams; heroin, 3 grams); if the amount of drugs is found to be below those limits, the prison sentence is at most six months or a fine or possibly even therapy, while a drug amount above the limits, depending upon the circumstances (e.g., member of criminal gang, repeat offender, head of a criminal organisation), can lead to much longer sentences.

On the other hand, courts in the United States not only pay particular attention to the difference between ‘possession’ and ‘trafficking’, but their drug laws and statutes are constructed so as to penalize more severely those who are caught with a larger quantity of illicit drugs than those caught with less. In fact, sentences are carefully graduated to correspond to increasing amounts of seized illicit drugs. In the United States, the total quantity  $Q$  of illicit drugs found in the possession of a defendant is a major determinant of the length of the sentence the court confers on that defendant. Hence, the value of  $Q$  becomes a very important component of the sentencing process.

In large-scale drug trafficking cases, where huge quantities of suspicious-looking substances are seized from defendants, the calculation of  $Q$  is tedious and expensive, and very demanding of forensic manpower, all of which can cause court delays (Beaupre and Eisler, 1996). To overcome this type of situation, a greater number of drug-testing laboratories have decided to use scientific sampling of the evidence which leads to statistical estimates of  $Q$ , and many jurisdictions are accepting such statistical procedures.

In a recent momentous decision, the U.S. Supreme Court in *Apprendi v. New Jersey* (2000), a case which had nothing to do with illicit drugs, established a rule that has since affected the prosecutions of all drug dealers and traffickers in the United States. *Apprendi* had a huge impact on the way federal prosecutors take their drug cases to court, especially how they estimate  $Q$  and present that estimate in court. The immediate effect of this decision was to encourage defendants convicted of drug charges to appeal their sentences. This led to hundreds of appeals of sentences, the vast majority of which were dismissed because defendants misinterpreted the meaning and scope of the *Apprendi* decision. All the ramifications of *Apprendi* have not yet been worked out and further clarifying decisions by the Supreme Court are still expected on those issues.

This article discusses two related situations: the manner in which  $Q$  was determined before *Apprendi* was decided, and the effect that *Apprendi* has had on determining  $Q$ . It is difficult

to understand the idiosyncracies of sampling practice in drug cases and the subsequent estimation of  $Q$  without first having some understanding of the legal context. In Section 2, we describe the legal context, including the salient differences between the U.S. federal courts and the state courts, the drug statutes and sentencing protocols, and two main kinds of standards of proof used in criminal trials. In Section 3, we introduce the idea that sampling drug seizures in order to estimate  $Q$  is desirable for many reasons, and give some background of the role that sampling plays in the judicial process in non-drug cases. In Section 4, we give details of the statistical and legal issues that are relevant when sampling drug seizures. In Section 5, we discuss a number of sampling procedures for estimating  $Q$ , including multistage sampling and composite sampling. Also included in Section 5 are frequentist and Bayesian discussions of how to determine the appropriate sample size to use in drug cases when some of the seized substances may not actually be illicit drugs. In Section 6, we briefly discuss the *Apprendi* case and the impact that the Supreme Court opinion has had on sampling and estimation of  $Q$ . Because there are a number of practical ramifications of *Apprendi* involving statistical issues that have not yet been decided by the Supreme Court, we have listed them as questions worth contemplating.

## 2. The Legal Context

### 2.1 *U.S. Federal and State Court Systems*

The United States employs a dual system of state and federal courts. Each of the 50 states (plus the District of Columbia and the U.S. territories) has its own hierarchical court system, which includes the major trial courts (for all cases involving criminal law and civil law), intermediate appellate courts (where an unfavourable trial court decision may be changed if based upon a reversible error), and supreme court (the court of last resort in the state).

The parallel federal court system, as established by the Constitution (Article III), the Judiciary Act of 1789, various Acts of Congress, and subsequent case law, has jurisdiction over disputes between states, between citizens of different states, and cases involving federal issues such as federal regulations, interstate commerce, and the interpretation of federal statutes. It consists of U.S. magistrate's courts, 94 U.S. district courts, 12 U.S. courts of appeal (usually called "circuit courts"), and the U.S. Supreme Court. Each state has between one and four U.S. district courts, with the number depending upon population density of the state. The district courts operate much like the state trial courts. Each circuit court has

jurisdiction over a specific geographical region and reviews cases which are appealed from the district courts of that region. The first eleven Circuits each include at least three states, while the 12th Circuit hears cases arising in the District of Columbia and has appellate jurisdiction over many federal government departments. The U.S. Supreme Court is the highest appellate court in the nation with nine permanent justices, each justice nominated by the President and confirmed by the Senate; it is also mandatory for the Supreme Court to accept an appeal when a lower federal court declares a federal law to be unconstitutional. The Supreme Court is likely to accept an appeal in which state laws or state constitutions are in question only if they conflict with the U.S. Constitution, laws, or treaties. In cases where federal law and state law contradict each other, the Supremacy clause of the Constitution (Article VI) establishes that federal law supercedes state law every time, except where the corresponding state laws are more stringent.

## 2.2 *U.S. Drug Statutes and Sentencing*

### 2.2.1 Federal Courts

Federal statutes and associated penalties for drug violations can be found in Title 21 of the United States Code, §841 *et seq.*, popularly known as the “Controlled Substances Act.” According to these statutes, controlled substances are divided into five different “schedules,” according to their potential for abuse. The two highest schedules are:

**Schedule I** Those drugs (or other substances), such as lysergic acid diethylamide (LSD), heroin, and marijuana, that have high potential for abuse, have no currently accepted medical use in treatment in the United States, and where there is a lack of an accepted safe dose for use of the drug under medical supervision.

**Schedule II** Those drugs, such as powder and crack cocaine, that have high potential for abuse, have a currently accepted medical use in treatment in the United States or a currently accepted medical use with severe restrictions, and where abuse of the drug may lead to severe psychological or physical dependence.

21 U.S.C. §841 consists of two subsections. Subsection (a) spells out that drug trafficking (the manufacture and distribution of controlled substances) is a prohibited activity, and subsection (b) lists the penalties for violating subsection (a). The statutory penalties (sentencing ranges and fines) for first-time convictions for trafficking in powder cocaine, crack cocaine, heroin, and LSD are set out in Table 1. Prior to the Supreme Court’s ruling in

*Apprendi v. New Jersey* (2000), federal courts had taken the position that subsection (b) was a sentencing provision *independent* of the substantive charges listed in subsection (a), and was relevant only after a defendant had been convicted of a crime under subsection (a). The Supreme Court in *Apprendi* rejected this interpretation. Before we discuss this case and its impact upon federal prosecutions, it is important, first, to understand how the system worked prior to *Apprendi*.

Federal sentencing courts need to juggle two different sets of penalties for all criminal activities. On the one hand, there are the federal mandatory minimum penalties, as created by Congress. On the other hand, there are the United States Sentencing Guidelines (*U.S.S.G.* or “Federal Sentencing Guidelines”), as created by the United States Sentencing Commission (1995b, §2D1.1(c)), which was established in 1985 by Congress as an independent agency within the judiciary. It is important that any sentence handed down by a federal court not contradict either of these two sets of penalties.

The Federal Sentencing Guidelines operate as follows. Sentences meted out for drug trafficking are taken from a two-way “Sentencing Table”, which has boiled down the essence of the crime into two characteristics: (1) an “offense level” of the crime (43 rows) and (2) the defendant’s prior criminal history (6 columns). The offense level of a crime is an adjusted version of a “base offense level”, or BOL, a number from 1 to 43 read off a “Drug Quantity Table”, and is a number that represents the total drug quantity  $Q$  seized from the defendant. The adjustment depends upon the circumstances of the crime. The biggest factor affecting the length of sentence is the total quantity  $Q$  of illicit drugs seized from the defendant, where the greater the amount, the higher the BOL, and the longer the sentence. Table 2, which lists a fragment of the Sentencing Table, shows the sentencing ranges according to the two factors of BOL and type of illicit drug (powder cocaine, crack cocaine, heroin, and LSD) for a first-time conviction with no aggravating or mitigating adjustments, and is clearly a lot more detailed than Table 1.

For example, the mandatory minimum penalties state that a first-time trafficker in cocaine, heroin, or LSD can be sentenced anywhere from five to twenty years. This sentence is doubled for second-time offenders, while for third-time offenders, the penalty is life imprisonment. The Federal Sentencing Guidelines give the court narrower boundaries within which to sentence that offender, depending upon  $Q$  and the offender’s profile for prior criminal activity. If the offender is caught with 50–150 grams of crack cocaine, this translates into a BOL of 32, and

with no prior record, the sentencing table gives a sentence of 121–151 months. Federal judges are required to follow the Federal Sentencing Guidelines, while staying within the broad sentencing boundaries provided by the mandatory minimum penalties. In fact, the United States Sentencing Commission (1995, Section 5G1.1) states that “a maximum sentence set by statute trumps a higher sentence set forth in the Guidelines.”

### 2.2.2 State Courts

Individual states have their own statutes and penalties for illegal drug possession and trafficking. According to a recent survey by the Bureau of Justice Assistance (1996), sixteen states had implemented, or were about to implement, sentencing guidelines similar to those of the federal government, and a further five states have guidelines under study. Most states, however, have not yet adopted sentencing guidelines, and several others have tried but failed to implement such guidelines. All states make use of some version of mandatory minimum sentencing laws, especially for drug trafficking. Whether following guidelines or not, penalties of different states can vary dramatically, and depend also on the year in which they were enacted; typically, they will differ from the federal penalties. In most states, fines are also applicable and differ according to the drug amounts seized.

## 2.3 *Standard of Proof Required*

### 2.3.1 General Principles

Whether the prosecution can convince a court that the defendant was engaged in drug trafficking and that an estimated quantity of those drugs was involved in the crime will surely depend upon the standard of proof used by that court. There are several standards of proof used by state and federal courts in the United States. The two main standards are “beyond a reasonable doubt”, used in criminal trials, and “preponderance of the evidence” (“more likely than not”), used in civil trials, both of which are inherently probabilistic in nature. Justice Harlan, in a concurring opinion to the U.S. Supreme Court decision in *In re Winship* (1970), noted that:

In a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what *probably* happened. The intensity of this belief — the degree to which a factfinder is convinced that a given act actually occurred — can, of course, vary. In this regard, a standard of proof represents an attempt to instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication. Although the phrases

‘preponderance of the evidence’ and ‘proof beyond a reasonable doubt’ are quantitatively imprecise, they do communicate to the finder of fact different notions concerning the degree of confidence he is expected to have in the correctness of his factual conclusions.

Because of the need to protect the presumption of innocence, weigh the relative moral consequences of an erroneous conviction and an erroneous acquittal, and compensate the defendant for the state’s disproportionate power, the Supreme Court in *Winship* held that all elements of a charged criminal offense must be proved beyond a reasonable doubt before the accused can be convicted. In *Patterson v. New York* (1977), the Supreme Court went further, saying that, subject to certain constitutional limits, state legislatures have the authority to define the ‘elements’ that constitute criminal conduct and can decide what constitutes a crime. It is this issue of defining the ‘elements of the offense’ in illicit drug cases that has differentiated the sentencing process in state courts from that in federal court.

### 2.3.2 Federal Courts

In federal court, a defendant who is arraigned on charges of possession with intent to distribute a controlled substance will most likely plead guilty rather than go to trial. Indeed, in 1995, 97.5 percent (751 out of 770) and 89.3 percent (12,652 out of 14,164) of all federal convictions for simple drug possession and for drug trafficking, respectively, were resolved by plea bargaining (United States Sentencing Commission, 1995a, table 17). By 1998, almost 60% of all federal prisoners were incarcerated for drug-related crimes (King and Klein, 2000).

The issue of whether  $Q$  is an essential element of the offense which must be proved beyond a reasonable doubt has not yet been specifically addressed by the Supreme Court. Prior to *Apprendi*, all but a few of the federal circuits held that factual determination of  $Q$  was not included under 21 U.S.C. §841 as an essential element of the offense and, hence, was not subject to the reasonable doubt standard; see, for example, *United States v. Lam Kwong-Wah* (1992) and *United States v. Royal* (1992) for the federal appellate history of this issue.

If a defendant either pleads guilty or is found guilty beyond a reasonable doubt after a federal trial, drug quantity becomes an issue only at sentencing. At that time, a probation officer usually prepares a presentence report (including a determination of  $Q$ ) on the defendant to be used by the court. Disputed facts at sentencing are established by a judge using the preponderance standard. Furthermore, judges are allowed to consider even non-charged criminal activity (termed “relevant conduct”) in computing a defendant’s sentence. This has always been a controvertial point; see, for example, Greenwald (1994). Prior to *Apprendi*,

the value of  $Q$  was an issue of fact which had to be established by a preponderance of the evidence solely by the sentencing judge.

### 2.3.3 State Courts

As far as this author is aware, all state courts consider the amount  $Q$  of illicit drugs seized from a defendant to be an essential element of the possession charge, especially when there is a lesser included offense for possessing a smaller amount. As such, that quantity must be proved beyond a reasonable doubt to a jury.

## 3. Determination of $Q$ by Sampling

### 3.1 *Arguments for Sampling Drugs*

An accurate determination of  $Q$  (and of the purity  $P$  of those drugs) for U.S. courts typically requires a tremendous amount of work. The effort — time, money, and manpower — needed to undertake a complete examination or “census” (testing and weighing) of all seized evidence is enormous, especially in major cities where high levels of drug dealing are prevalent. See Izenman (2001) for examples of current practice in New York City and Orange County, California. The huge caseloads of the crime laboratories, whose job it is to examine drug seizures and determine  $Q$ ,  $P$ , and the types of drugs, make for court delays until examinations are completed.

Carrying out a complete census has a number of disadvantages, including:

1. When  $N$  is large, or when there are many drug seizures that need to be investigated, inaccuracies inevitably occur in testing and weighing because of time and manpower constraints, and so exact results cannot be guaranteed.
2. Pressure to report results in a short time to meet court scheduling gives prime opportunities for errors in testing, weighing, and counting.
3. Certain chemical testing unavoidably destroys the evidence, so that in a complete census, no drugs would remain to present to the jury, or for the defense to carry out its own testing if desired.
4. Testing the contents of all seized packages exposes forensic chemists to potential health hazards through airborne dust or physical contact (Le, Taylor, Vidal, Lovas, and Ting, 1992).

These arguments provide excellent reasons for randomly sampling the seized evidence and thereby estimating  $Q$ . It would certainly be beneficial for the courts to see prosecutors presenting drug cases using a moderately-sized sample that has been carefully tested for drugs, weighed, and counted.

Thus, U.S. forensic laboratories, both state and federal, often use scientific sampling of the evidence to present statistical estimates of  $Q$  and  $P$ . Sampling plans include simple random sampling, stratified sampling, multistage sampling, and composite sampling. The type of sampling adopted in any particular case is usually guided by the complexity of the seizure. This involves consideration of the number of containers (plastic bags, envelopes, bottles) seized, the nature of their contents (pills, tablets, capsules, vials, rocks, loose powder, liquid), and the types of drugs involved.

Early uses of sampling in drug cases were typically ad hoc. Sample sizes were conveniently chosen by the chemist and no information was given to the court as to whether or not the sampling was carried out randomly. Appeals of subsequent convictions often depended less on the randomness of the sampling mechanism than on legal issues regarding measurement procedures. Although a wide variety of sampling methods are used today in drug cases, the courts must decide whether a particular procedure is appropriate for the case in question.

### *3.2 Sampling in Non-Drug Cases*

The practice of sampling has long been accepted by the courts in a wide variety of criminal and civil cases. Larsen (1964) annotates over a hundred civil cases where “samples” were submitted in evidence of actions for recovery of damages by breaches of contract, warranty or covenant, and by negligence and personal injury or death. More recently, Walker and Monahan (1999) studied the use of sampling to prove causation in lawsuits brought by states seeking reimbursement from tobacco companies for Medicaid payments attributable to tobacco-related diseases. The results of specially-commissioned surveys and opinion polls have also been used to help decide damages in trademark infringement cases, antitrust cases, accounting cases, and motions for a change in venue in criminal cases; see Diamond (1994), Strong (1992, Section 208), and Walker and Monahan (1998). Surveys have been accepted by the courts in cases involving false advertising and consumer confusion about the effectiveness of certain products; see, for example, *Novartis Consumer Health, Inc. v. Johnson & Johnson-Merck Consumer Pharms. Co.* (2002). Sampling has also been utilized as a means for determining damages in cases of copyright infringement and software piracy, trafficking

in cyberporn, and the waste, fraud, and abuse of U.S. federal welfare programs; see Izenman (2000a). For an excellent discussion of samples used in evidence, see Gastwirth (1988, Chapter 9).

Courts have been very accepting of sample data even if there exist alternative methods of proof which may be difficult to carry out. In many instances, samples have been introduced into evidence to establish a particular point and to illustrate the condition, quality, or nature of a large amount of material which is involved in the matter under litigation but is not accessible to the court. Wigmore (1979, Section 439) gave two requirements for a sample, presented to show the quality or condition of the entire lot or mass from which it was taken, to be admissible in evidence: (1) the mass should be substantially uniform with reference to the quality in question (a “homogeneity” requirement) and (2) the sample portion should be of such a nature as to be fairly representative (see Kruskal and Mosteller, 1979).

## 4. Sampling Issues for the Courts

We now look at the sampling issues raised by those states that have been most deliberative and cautious when it comes to accepting an estimate of the total quantity of illicit drugs seized from a defendant. If convicted, the defendant typically challenges the sufficiency of the evidence to support conviction on trafficking charges, arguing that the prosecution failed to prove beyond a reasonable doubt that he or she possessed more than the statutory threshold amount of a substance containing the illicit drug. Usually, at issue is the legal weight to give the results of random sampling in such cases because not all seized items are tested. The conviction and sentence of the trial court will usually be affirmed by a higher court if the method of sampling allows the court to accept that the untested items also contain illicit drugs and, hence, provides a quantity estimate that satisfies the required standard of proof.

We divide up the following discussion (corresponding to the divisions between the courts) into situations in which the drugs are found in a single container or in multiple containers.

### *4.1 Drugs Found in a Single Container*

When it comes to the seizure of a single container, the typical situation is that the drugs found inside that container take the form of pills, tablets, capsules, or rocks. There may be hundreds or thousands (if not tens of thousands) of such items in the container and a complete inspection of every item would clearly be impracticable. In general, if the items

found inside a single container can be regarded as homogeneous (through visual appearance or preliminary chemical testing of several items), then courts have accepted an examination by a forensic chemist of a random sample of those items to prove the identity of the remainder of the items in that container. For example, in *People v. Kaludis* (1986), an Illinois Court of Appeals affirmed the conviction in which a chemist randomly selected three tablets from a bag containing 100 homogeneous tablets found in the defendant's possession, tested the three tablets positively for methaqualone, and concluded that all 100 tablets contained methaqualone.

#### 4.2 *Drugs Found in Multiple Containers*

In the event that the suspected contraband (usually a powdery substance) is found in multiple bags, containers, or receptacles, certain jurisdictions have specifically distinguished the sampling situation from that of sampling similar-looking pills in a single container.

##### 4.2.1 State Courts.

The first specific sampling policy for multiple containers was formulated by the Illinois appellate courts, which addressed the issue of how to handle any questionable homogeneity of the various containers and their contents. It is important to understand this policy because it may have repercussions for future drug trials in the federal courts.

A general rule that guards against possible heterogeneity of the evidence was developed consisting of three major points. This rule can be expressed as follows. When a defendant can be charged with the lesser-included offense of possession of a smaller amount of an illicit drug, the weight  $Q$  of the seized drug is an essential element of the crime and must be proved beyond a reasonable doubt. In such a situation:

1. At least a sample from *each* container must be conclusively tested to prove that it contains an illicit drug;
2. the contents (and weight) of the untested containers may not be considered in determining the severity of the offense; and
3. if the total weight of the sampled containers that conclusively tested positive for an illicit drug exceeds the minimum amount needed to prove the charge against the defendant, then (1) can be ignored.

In essence, the courts regarded the containers as strata, and the sampling procedure should, therefore, conform to the classical notion of stratified random sampling whereby a random sample of the contents is taken from each stratum. We shall henceforth find it convenient for exposition purposes to refer to this as the “Illinois stratified sampling rule.”

This sampling policy was originally derived from a series of three cases: *People v. Games* (1981), *People v. Ayala* (1981), *People v. Hill* (1988). In each of these cases, the defendants had been convicted at a jury trial of possession of more than 30 grams of cannabis, heroin, and cocaine, respectively. In the appeals of the *Games* and *Ayala* cases, the courts refused to buy the State’s extrapolations to two seized bags when only one bag had been conclusively tested and weighed; they ruled that when there are two containers, both must be analyzed to be admitted as evidence. The *Hill* case generalized this ruling to more than two containers. In each of these three cases, the sentence was reduced to a lesser-included offense of possession of less than 30 grams of an illicit drug.

This should have settled the issue. However, in *People v. Green* (1993), the Illinois appellate court affirmed a conviction based upon the contents of a sample of the containers seized from the defendant. The court in *People v. Black* (1994) then formally extended the scope of the *Kaludis* rule so that multiple “similar” containers could be treated for sampling purposes as if they were pills or capsules. This situation was not to last long. In *People v. Jones* (1996), the appellate court reinstated the stratified sampling rule, noting that “although we agree with the holding in *Kaludis*, we find the extension set forth in *Black* to be overly broad . . . We believe the opinion in *Black* represents an unwarranted departure from the long-standing rule requiring the State to test an adequate number of samples with a sufficient combined weight to establish the elements of the offense”; as in *Hill*, the conviction was reduced to a lesser-included offense. The State appealed *Jones* to the Illinois Supreme Court, whose majority opinion offered no new insight or principle, but just repeated the stratified sampling rule in affirming the appellate court’s reduction in the defendant’s sentence. In a dissent, two of the justices (including the Chief Justice) stated that

There is no reason to require the State in cases such as this to test the contents of each of the items the defendant has in his possession. Random sampling can provide circumstantial evidence of guilt, the strength of which will vary from case to case. Today’s decision simply imposes an unnecessary burden on the State, making more difficult the prosecution of offenders who are found with contraband divided among multiple bags, packets, or other containers that, under the majority’s rule, must now be tested individually.

Although many other jurisdictions have considered the question of multistage sampling, only Florida and Minnesota adopted the Illinois stratified sampling rule, reducing convictions and sentences where the stratified sampling rule was not followed. Other State jurisdictions give no special distinction to sampling from multiple containers and generally affirm two-stage sampling procedures in drug cases.

#### 4.2.2 Federal Courts.

Prior to *Apprendi*, all federal courts accepted multistage sampling because the standard of proof for sentencing was preponderance of the evidence rather than beyond a reasonable doubt.

## 5. Estimating the Amount of Drugs

### 5.1 Drug Quantity $Q$

Suppose a defendant is found in possession of  $K \geq 1$  containers (such as plastic bags, ziplock “baggies,” glassines, paper packets or envelopes, vials, or even bottles). These  $K$  containers may be packaged into, say,  $B \geq 1$  batches, where the  $b$ th batch consists of  $K_b$  containers and  $K = \sum_{b=1}^B K_b$ . For example, in *People v. Argro* (1975),  $B = 10$  and  $K = 148$ , and in *State of North Carolina v. Hayes* (1976),  $B = 2$  and  $K = 19$ . For convenience and because most seizures consist of only one batch, we assume that  $B = 1$ .

Suppose the weight of the  $t$ th item in the  $j$ th container is recorded as  $X_{jt}$  grams if the item tests positive, while if that item tests negative the weight is recorded as 0 grams. Then, the total weight of the drugs seized is

$$Q = \sum_{j=1}^K \sum_{t=1}^{N_j} X_{jt}, \quad (1)$$

where  $N_j$  is the number of items in the  $j$ th container,  $j = 1, 2, \dots, K$ . If  $Q$  falls into a specific range or interval,  $I = [a, b)$ , say, given in the statutes, then the penalty is determined as so many years of imprisonment plus a fine of a certain amount.

Thus, it is important to determine as accurately as possible the value of  $Q$  and the purity  $P$  (as a percentage) of those drugs. In New York (see *People v. McLaurin*, 1993), it is total “pure weight”,  $Q \times P$ , which gets reported to the court and determines the defendant’s responsibility. If the drugs seized are perforated sheets of 100 “one-dose” squares of a carrier medium (usually blotter paper) impregnated with LSD for oral ingestion, then the problem is

to determine how many of those squares contain LSD, a number which in some jurisdictions, is then converted into a total weight measurement  $Q$ .<sup>1</sup>

## 5.2 *Sampling for Illicit Drugs*

A much more practical procedure than carrying out a complete examination of all seized evidence would involve randomly sampling the seized evidence and thereby estimating  $Q$ . An important point to make here about sampling inanimate items (such as pills, capsules, envelopes) is that if we wish to take a simple random sample of the items, they have to be homogeneous in appearance. Homogeneity is a concept that is linked directly to simple random sampling. If every item in the seizure were tested there would be no need to worry about homogeneity of the items. If all items in a package appear alike due to visual inspection of their shape, color, and markings, then certain courts have held that they are indeed homogeneous and evidence based upon sampling those items is accepted. Arguments involving the homogeneity issue have been put forward and accepted for tablets, pills, capsules, rocks and bricks of crack cocaine, and LSD on blotter paper (Izenman, 2001).

A referee has made the point that while a sample can be randomly drawn from a well-defined population (regardless of whether the items are homogeneous in appearance or not) and a valid statistical estimate of  $Q$  can in turn be provided, it will probably be more difficult to convince the court of the validity of such an estimate of  $Q$  if the homogeneity issue has been ignored in the sampling process.

The situation is somewhat different when it comes to loose powder wrapped in some packaging medium, such as small plastic bags (e.g., Ziploc “baggies”), and where no visual clues of homogeneity can be readily discerned. In such cases, it is necessary to carry out a more careful testing of the evidence. There have been many cases which have been reversed solely because of negligence and a lack of sufficient testing needed to convince the court or the sentencing judge. The main reason why extra care is required in testing loose powder is due to the fact that powder – white, brown, or any other color — may not always be an illicit substance. Courts have recognized that drug dealers are known to sell “placebo” or counterfeit drugs, which may be milk sugar, flour, baking soda, powdered sugar, pancake mix,

---

<sup>1</sup>In Britain, purity usually becomes a sentencing issue in cases of drug smuggling or where at least 500 grams of cocaine, heroin, or amphetamine powder are seized. In such cases, weight is based on the  $Q \times P$  formula. In the specific cases of Ecstasy tablets or LSD in dosage form, the amount is assessed by reference to the number of tablets or doses, with an assumed average purity of 100 mg of Ecstasy and 50 mg of LSD, unless prosecution or defense, by expert witness, shows otherwise. See *R v. Morris* (2000).

or soap pieces. As a result, many states (and the U.S. government under 21 U.S.C. 841(a)(2)) have laws against distribution of counterfeit drugs that are made up to resemble actual drugs.

### 5.3 Drug Smuggling

It is easy to argue homogeneity when prosecuting smuggling enterprises that import illicit drugs into the United States. Many of these drugs are brought into the United States through couriers or “mules” who do so by swallowing large numbers of balloons (a generic term, usually referring to condoms, prophylactics, capsules, or pellets) filled with heroin or cocaine. These balloons remain in the alimentary canal of the smuggler until it is time to expel them. The smugglers are usually blackmailed into carrying out the trips by smuggling organizations, who then teach their mules how to swallow over a hundred balloons each trip. Many of these smugglers get caught when they arrive in the United States. The general consensus is that no-one carries counterfeit drugs in balloons from these countries to the United States and that it is highly likely that the drugs that are found inside the stomach of a smuggler are from the same batch of drugs with roughly the same purity.

For more specific descriptions of drug smuggling techniques and the related statistical problems (including the effect of measurement error) involving estimation of  $Q$ , see Izenman (2000*b, c*).

### 5.4 Multistage Sampling and Estimation of $Q$

It will be convenient to denote the set of containers by  $\mathcal{K} = \{1, 2, \dots, K\}$ , and the set of items within the  $j$ th container by  $\mathcal{N}_j = \{1, 2, \dots, N_j\}$ ,  $j = 1, 2, \dots, K$ . We are led to consider the following two-stage sampling design:

- (I) A simple random subsample,  $\mathcal{K}_0 \subseteq \mathcal{K}$ , of size  $k$  is drawn from the  $K$  containers. The probability of choosing the  $j$ th container is  $\pi_1 = k/K$ ,  $j \in \mathcal{K}_0$ .
- (II) From each container  $j \in \mathcal{K}_0$  selected at the first stage, a simple random subsample,  $\mathcal{N}_{j0} \subseteq \mathcal{N}_j$ , of size  $n_j$  is drawn from the  $N_j$  items contained therein, where subsampling of one container is independent of subsampling of all other containers. The probability of choosing the  $t$ th item from the  $j$ th container is  $\pi_{2j} = n_j/N_j$ ,  $j \in \mathcal{K}_0, t \in \mathcal{N}_{j0}$ . Each of the  $n_j$  sampled items is then tested for the presence of an illicit drug.

Note that we allow the possibility that all the containers may be chosen. So, in general, we wish to select containers  $\{\mathcal{K}_0\}$  and then items  $\{\mathcal{N}_{j0}\}$  such that  $n = \sum_{j \in \mathcal{K}_0} n_j$  is the total

number of sample items. Determinations of  $k$  and  $n$  that have been used in the courts are discussed below.

Suppose  $m_j$  out of the  $n_j$  items sampled from container  $j \in \mathcal{K}_0$  test positive for a controlled substance. If  $m_j = 0$  for any  $j$ , we say that the  $j$ th sample container tests negative for an illicit drug. Let the number of sample containers testing positive be denoted by  $k^+$ . As before, let  $X_{jt}$  be the weight of the  $t$ th sample item if that item tests positive, and set  $X_{jt} = 0$  if it tests negative,  $j \in \mathcal{K}_0, t \in \mathcal{N}_{j0}$ . Then, the two-stage Horwitz-Thompson estimator,

$$\widehat{Q}_{2st} = \sum_{j \in \mathcal{K}_0} \sum_{t \in \mathcal{N}_{j0}} \frac{X_{jt}}{\pi_1 \pi_{2j}} = \frac{K}{k} \sum_{j \in \mathcal{K}_0} \frac{N_j}{n_j} \sum_{t \in \mathcal{N}_{j0}} X_{jt}, \quad (.2)$$

based upon  $m^+ = \sum_{j \in \mathcal{K}_0} m_j$  sample items testing positive, is unbiased for  $Q$  in (1). It is not difficult to show (Sarndal, Swensson, and Wretman, 1992, chap. 4) that an unbiased estimate of the variance of (2) is given by

$$\widehat{\text{var}}(\widehat{Q}_{2st}) = K^2 \left( \frac{1 - \pi_1}{k} \right) s_1^2 + \frac{K}{k} \sum_{j=1}^K N_j^2 \left( \frac{1 - \pi_{2j}}{n_j} \right) s_{2j}^2, \quad (.3)$$

where  $s_1^2 = (k-1)^{-1} \sum_{j \in \mathcal{K}} (y_j - \bar{y})^2$ ,  $s_{2j}^2 = (n_j-1)^{-1} \sum_{t \in \mathcal{N}_j} (X_{jt} - \bar{x}_j)^2$ ,  $y_j = (N_j/n_j) \sum_{t \in \mathcal{N}_j} X_{jt}$ ,  $\bar{y} = k^{-1} \sum_{j \in \mathcal{K}} y_j$ , and  $\bar{x}_j = n_j^{-1} \sum_{t \in \mathcal{N}_j} X_{jt}$ . Similar expressions for three-stage estimators (assuming containers are randomly sampled from batches) and estimators of their variances can also be obtained.

### 5.5 Composite Sampling

Many crime laboratories are now using composite sampling of the drug-filled containers in order to estimate  $Q$  and the purity  $P$ . This sampling technique cuts down on the high costs incurred by the laboratories in analyzing the contents of possibly thousands of seized containers. We refer the reader to the article by Lancaster and Keller-McNulty (1998), who provide an excellent review of composite sampling, and to the annotated bibliography by Boswell, Gore, Lovison, and Patil (1996).

In the case of drug seizures, a composite sample is obtained by a forensic chemist as follows; see, for example, the New York State case *People v. McLaurin* (1993). A random sample is taken of  $k$  out of the  $K$  containers, removing  $c$  core samples from the contents of each selected container, putting the core samples into a mortar and then grounding them up with a pestle into a very fine powder, which becomes the composite sample. The powder is mixed by repeated tossing and stirring to look as homogeneous as possible. Using a spatula,

$R$  subsamples are randomly scooped up from around the mixture and placed into weighing dishes, where they are tested for the presence of an illicit drug. Usually,  $c$  is taken to be 4 or 5, and  $R$  is taken to be 1 or 2. If a drug is present in at least one of the core samples prior to mixing, then a subsample from that mixture will, with high probability, test positive for that drug. On the other hand, if the subsample tests negative for that drug, then no further sampling or testing would be carried out, and all containers would be declared free of that drug.

Composite sampling in drug cases takes various forms. On the one hand, some forensic laboratories combine *all* of the contents of each container before sampling and testing; see, for example, *Mello v. State of Texas* (1991), *People v. Little* (1986), *People v. Jackson* (1985), and *State of North Carolina v. Clark* (1973). Although this strategy certainly satisfies the stratified sampling rule, it leaves open the questions of exactly which and how many of the containers possess drugs if the mixture tests positive. A more reasonable strategy would be to mix portions of the contents of each container; see, for example, *Pugh v. State of Georgia* (1995).

It is not uncommon for forensic chemists to misinterpret the results derived from composite sampling. In particular, testimony is often given that illicit drugs are present in *all*  $K$  containers when only a single subsample ( $R = 1$ ) from a composite sample tests positive. In fact, a positive test result implies only that at least one of the core samples contains an illicit drug. We also see appeals courts focussing on the sampling part of the composite sampling method rather than the inferential part. If core samples were not taken from every one of the seized containers ( $k < K$ ), and if the jurisdiction is one in which the Illinois stratified sampling rule operates, then the court will not accept the estimate of  $Q$ . If every container provides a core sample for the composite sample ( $k = K$ ), then that same court would accept the poor inference that there are drugs in all  $K$  containers.

As a result, composite sampling in forensic laboratories has attained a controversial status. A dissenting opinion in *People v. Little* (1986) completely disagreed with the practice of composite sampling and remarked that “I cannot believe that the majority here would have allowed the police to combine one of those manila envelopes with, say, a five-pound bag of flour or sugar, had such an item also been in the defendant’s possession. That, however, is where their logic leads them”. We note that when an illicit drug is present amongst the substances within a container and a core sample from that container is mixed with core

samples from one or more containers that do not contain drugs (usually called “adulterants”), the resulting composite sample will test positive for that drug, but it will be the purity measurement  $P$  that will be affected: the greater the proportion of adulterants, the lower the purity. See, for example, *Ross v. State of Florida* (1988). Until uniform guidelines for sampling, testing, and estimation are implemented when dealing with composite sampling, such controversial issues will not be resolved.

### 5.6 Sample Size

When sampling from one or more containers, the choice of sample size is important. If all the items in a container appear to be visually homogeneous in the sense described in Section 5.3, then we can use the standard formulas for sample size determination as can be found in Cochran (1977, Chapter 4) or Desu and Raghavarao (1990). An example (see Section 5.4) is international drug smuggling.

One unusual practice of determining sample size has been what Izenman (2001) calls the *square-root rule*. This rule, which has been used since the 1920’s by the Association of Official Agricultural Chemists (operating with the blessing of the U.S. Department of Agriculture), states that if the number of items in the container is  $N$ , then the sample size should be taken to be  $n = \sqrt{N}$ . This rule has no statistical justification and was started as a convenient rule-of-thumb for agricultural regulatory inspectors in the field who knew how to extract a square root. Unfortunately, the square-root rule is still being used today for drug sampling by federal regulatory agencies (such as the Drug Enforcement Administration), state agencies, and forensic laboratories. A recent international study by Colon, Rodriguez, and Diaz (1993) found that the square-root rule was the most popular rule in Australia, Austria, Canada, England, New Zealand, Hong Kong, Northern Ireland, the United States, and the U.S.A. Army-Europe, for deciding how many containers or items, whether homogeneous or not, to select in a drug testing situation. Furthermore, the United Nations (1987) recommends the square-root rule for drug testing when using composite sampling. The amazing popularity of the square-root rule is a prime example of how an unfounded rule-of-thumb retains its importance in practice even though it is theoretically unjustified.

The square-root rule is not the only theoretically-unjustified rule that is used today for determining sample size for drug testing. The study by Colon, Rodriguez, and Diaz (1993) also found that a 10% rule for sample size (which may have originated with Deming, 1954) is being used by forensic laboratories in Australia, Canada, and the United States. A 4%

rule is used in England, half the square-root rule is used in Switzerland, and in the United States, some laboratories use a fixed sample size rule, such as 1, 4, or 15 containers or items, regardless of the amount of the contraband seized.<sup>2</sup>

It is not uncommon, however, for the contents of a container not to be obviously homogeneous and filled with drugs. Indeed, there may be concern that not all of the contents of the containers are actually drugs. This situation occurs most frequently when the material seized is loose powder, which could be heroin or cocaine, for example, or possibly a counterfeit drug, such as flour, sugar, or pancake mix. The counterfeit drugs could be made of harder material, such as pieces of soap or baking soda that are designed to look like rocks of crack cocaine. In such cases, to comply with sentencing thresholds, we would expect the chemist to take this uncertainty into consideration when choosing the sample size.

There have been both frequentist and Bayesian approaches to sample size determination that take into account the possible presence of counterfeit drugs within the seized containers. Some of those who have worked on this problem include Frank, Hinkley, and Hoffman (1991), Tzidony and Ravreby (1992), and Aitken (1999, 2000).

Let  $N < 50$  be the number of seized items (or containers). Suppose  $N = N_0 + N_1$ , where  $N_0$  of the items contain no drugs while  $N_1$  contain drugs. We wish to take a random sample of size  $n$  from the  $N$ . Let  $Y$  be the observed number of items in the sample that do not contain drugs. Then,  $Y$  has a hypergeometric distribution,

$$\text{Prob}[Y = y | N_0, N_1, n] = \frac{\binom{N_0}{y} \binom{N_1}{n-y}}{\binom{N}{n}}, \quad (.4)$$

where  $\max\{0, n - N_1\} \leq y \leq \min\{n, N_0\}$ . We often wish to test  $N_1 < k$  against  $N_1 \geq k$ , where  $k$  is some given value chosen by the forensic chemist. For example,  $k$  might represent the smallest number of items that contain drugs (out of the  $N$  seized) for the defendant to be sentenced to a specific term of imprisonment. We can set  $k$  to be an unknown proportion of  $N$ ,  $k = \theta_0 N$ , say, where  $0 < \theta_0 < 1$ . Then, based upon the sampling evidence and with given probability  $1 - \alpha$  ( $0 < \alpha < 1$ ), we would like to reject the possibility that  $N_1 < k$  for some prespecified  $k$ . If  $\alpha = 0.05$  and if we set  $\theta_0 = 0.98$ , the problem boils down to determining a sample size  $n$  such that with 95% confidence at least 98% of the  $N$  items contain drugs. Clearly, we would reject  $N_0 < k$  for the hypergeometric distribution if  $Y$  is too small. Let  $m_0$

---

<sup>2</sup>In non-drug-related sampling, we note that British Standard BS 5309 (1993) uses a sample size of  $n = 3 \times N^{1/3}$  to sample chemical products.

be the largest number of items in the sample that the forensic scientist would expect to test negative for drugs. Frank, Hinkley, and Hoffman (1991) showed that the required sample size  $n$  can be obtained by solving the inequality,

$$\sum_{i=n-m_0}^n \frac{\binom{k-1}{i} \binom{N-k+1}{n-i}}{\binom{N}{n}} \leq \alpha, \quad 0 < \alpha < 1. \quad (.5)$$

If we set  $m_0 = 0$ , then the inequality reduces to

$$\frac{(k-1)!(N-n)!}{N!(k-n-1)!} \leq \alpha, \quad (.6)$$

which, for given values of  $N$ ,  $k$ , and  $\alpha$ , can be solved for  $n$ . The more general inequality can be solved for  $n$  by using an iterative algorithm that would loop through possible values of  $k$  and  $n$  until it finds the largest value of  $k$  and the smallest value of  $n$  to satisfy the inequality. Values of  $n$  that satisfy the inequality for  $\alpha = 0.05$  and  $N = 150$  with  $m = 0(1)10$  are given in Hedayat, Izenman, and Zhang (1996, Table 2). For example, if we have  $N = 150$  seized items and we wished to claim with 95% confidence ( $\alpha = 0.05$ ) that the number of items containing drugs is  $N_1 \geq k = 0.98N = 147$ , while expecting no negative items in the sample ( $m_0 = 0$ ), then a sample size of  $n = 79$  would solve the above inequality. If we allow a single negative item in the sample ( $m_0 = 1$ ), the value of  $n$  escalates to 113, while allowing two negative items ( $m_0 = 2$ ) yields  $n = 135$ . If we tone down our expectations and set  $\theta_0 = 0.8$  in the above calculations, then we would have  $k = 120$ ; with no negative item expected, the sample size would be  $n = 13$ ; one negative would yield  $n = 20$ , and two negatives  $n = 27$ . See also Frank, Hinkley, and Hoffman (1991, Table 2).

A Bayesian approach to the problem of determining sample size when we allow for the possibility that negative items may be mixed into the seized items has also been studied. Assume that  $N$  is large enough that sampling can be regarded as being with replacement. In this instance,  $Y \sim \text{Bin}(n, 1 - \theta)$  is binomially distributed with  $\theta$  equal to the probability that an individual item contains drugs. If we take a Beta( $a, b$ ) conjugate prior density for  $\theta$ ,  $0 < \theta < 1$ ,  $a, b > 0$ , then the posterior density is a Beta( $n - y + a, y + b$ ) density. We wish to find  $n$  such that we will be  $100(1 - \alpha)\%$  confident that at least  $100\theta_0\%$  of the seized items contain drugs ( $0 < \theta_0 < 1$ ) if all  $n$  items test positive for drugs (i.e.,  $y = 0$ ). This gives us an integral equation,

$$[B(n + a, b)]^{-1} \int_{\theta_0}^1 \theta^{n+a-1} (1 - \theta)^{b-1} d\theta = 1 - \alpha, \quad (.7)$$

that is solved for  $n$  using a trial-and-error strategy. For examples of this approach, see Aitken (2000).

## 6. *Apprendi* and its Effect on Drug Prosecutions

### 6.1 *Apprendi v. New Jersey (2000)*

The facts of this case are deceptively simple and on the surface appear to have little or nothing to do with estimating  $Q$  in drug cases. On 22 December 1994, a white pharmacist, Charles C. Apprendi, Jr., living in Vineland, New Jersey, apparently angry that a black family had moved into his neighborhood, fired several shots into their home. No-one was harmed in the incident, but Apprendi pleaded guilty to possession of a firearm for an unlawful purpose, which was a second-degree felony under New Jersey law and punishable by a term of 5–10 years in prison. The charge against Apprendi did not include any reference to the State’s hate crime statute, which enhances a sentence if a trial judge finds, by a preponderance of the evidence, that the defendant committed the crime to intimidate a person or group because of his or her race. The sentencing judge, by a preponderance of the evidence, then found the crime to be racially motivated, triggering New Jersey’s hate crimes statute, which, in turn, increased the sentencing range to 10–20 years. Apprendi was consequently sentenced to 12 years in prison. In June 2000, a divided (5–4) Supreme Court reversed that sentence, striking down New Jersey’s hate crime law as unconstitutional, and announced a constitutionally-based rule:

Other than the fact of a prior conviction, any fact that increases the penalty for a crime beyond the prescribed statutory maximum must be submitted to a jury, and proved beyond a reasonable doubt.

Apprendi’s sentence was later reduced to seven years, with a minimum of three years to be served; he actually served five years.

Justice O’Connor, in dissent (and joined by Chief Justice Rehnquist and Justices Kennedy and Breyer), characterized the ruling as a “watershed change in constitutional law” in the sense that it would significantly change the course of criminal litigation. She also predicted (accurately as it turns out) that “the number of individual sentences drawn into question by the Court’s decision could be colossal.” She also noted that the majority were “[casting] sentencing . . . into what will likely prove to be a lengthy period of considerable confusion.” Although this case did not involve drugs, legal commentators at that time swiftly concluded that the ruling could affect all criminal prosecutions, especially the constitutionality of drug

sentencing under 21 U.S.C. §841(b), and warned that the entire Federal Sentencing Guidelines could be undermined. To get an idea of the magnitude of this case, a LEXIS-NEXIS search revealed that after only two and a half years, there had been over 4000 cases citing the *Apprendi* decision.

## 6.2 *The Impact of Apprendi on Drug Sampling*

The surprising *Apprendi* decision “sent shock waves through the criminal justice system” (Greenhouse, 2001). After an initial period of judicial chaos and confusion as to the practical effect of the *Apprendi* decision, prosecutors were essentially forced to change the way they prepared and presented their drug cases at trial. The immediate effect of *Apprendi* was that defendants convicted in drug cases prior to *Apprendi* interpreted their convictions and sentencing as having violated the rule announced in *Apprendi*. This unleashed a flood of appeals, the vast majority of which were dismissed because they did not involve sentences that extended past the statutory maximum.

As a result of the *Apprendi* ruling, prosecutors are now being extra-cautious in drug cases by making sure that indictments declare every element of the offense, and that evidence is presented about the exact amount of drugs and the amount of money alleged to be involved in the crime. Almost all federal circuits now agree, post-*Apprendi*, that under 21 U.S.C. §841 the specific drug amount  $Q$  has to be considered an element of the offense and stated explicitly in the indictment (only the Seventh Circuit has disagreed). However, the federal courts have since also concluded that failure to do so may not be fatal.

The *Apprendi* decision does not change the law that a judge, by a preponderance of the evidence, can determine facts, such as drug type and quantity, related to a defendant’s sentence — as long as the resulting sentence stays within the boundaries of a range with a maximum of 20 years as provided by 21 U.S.C. §841(b)(1)(C). For example, *Apprendi* is irrelevant if a convicted drug trafficker dealing in Schedule I or II controlled substances is sentenced to less than 20 years in prison. The *Apprendi* decision also specifically refused to reverse an earlier opinion whereby a judge could determine, by a preponderance of the evidence, whether an enhancement could apply, and instead limited that case’s holding to cases whose sentences do not violate the statutory range on the conviction as determined by the jury. Similarly, *Apprendi* is irrelevant when the statutory maximum is life imprisonment. If, however, the factors in the case can lead to a sentence greater than the statutory maximum,

then those factors now have to be specified in the indictment and proved by the reasonable doubt standard to a jury.

If a determination of  $Q$  (or an estimate  $\hat{Q}$ ) is included in the indictment, the jury has to consider its veracity, given the evidence, beyond a reasonable doubt. Previously, the value of  $Q$  was not included in the indictment, and  $Q$  was decided by a judge using the preponderance-of-the-evidence standard of proof. Because the latter standard is more lenient than the former, the following questions regarding sampling suggest themselves.

1. In an indictment, how should the prosecution charge a defendant of trafficking in a quantity  $Q$  of drugs, especially if  $Q$  has to be estimated? *Apprendi* may not necessarily require an exact drug quantity, 223.8 grams, say, to be charged in the indictment and determined by a jury. It may be sufficient instead to have the jury find for a certain quantity range, 200–300 grams, say, for powder cocaine that would result in a sentencing range, 33–41 months, say, for a first offender as specified in the Federal Sentencing Guidelines (see Table 2).

The range of  $Q$ -values could be derived from an approximate 95% confidence interval,  $\hat{Q} \pm 2 \cdot \hat{\text{se}}(\hat{Q})$ , where  $\hat{\text{se}}(\hat{Q})$  is the estimated standard error of  $\hat{Q}$  (see Section 5.5). Most likely, the confidence interval will lie wholly within a range of  $Q$ -values in Table 2, and so that range can be stated in the indictment. On the other hand, if the confidence interval lies partly in one range and partly in an adjacent range (as when  $Q$  (or  $\hat{Q}$ ) is determined to have a value very close to a boundary in the Sentencing Table), then we have an issue that would have to be determined by the jury. One way of addressing this issue before it gets to court would be to increase the sample size  $n$  in the sampling and inference process. This would reduce the estimated standard error and the width of the approximate confidence interval, and, hence, also reduce the likelihood of the approximate confidence interval crossing a sentencing boundary. However, on this point, see (3) below.

As statistical guidance to a jury, it may be necessary for the prosecution to provide a statistician as expert witness to explain the general concept of sampling error and, hence, of the importance and use in sentencing of an estimated standard error of  $\hat{Q}$ .

2. Would the forensic laboratories have to change their sampling and estimation methods now that a jury is required to decide on the value of  $Q$  using a much tougher standard

(beyond a reasonable doubt) than a sentencing judge would need to use (preponderance of the evidence)? Should we expect the federal circuits to revise the sampling protocols that are acceptable in their courts for drug sampling and estimation of  $Q$  to something like the Illinois stratified sampling rule (see Section 4.2.1)?

3. If the courts find that multistage and composite sampling are still acceptable sampling protocols, would larger sample sizes than before be required to persuade the jury beyond a reasonable doubt? This has been a point of some contention amongst certain state police departments who have long believed that the more evidence the jury sees and hears about, the more likely a conviction can be obtained under a state's beyond-a-reasonable-doubt standard. This rationale has been used by state police departments to push for a sample closer to a complete census (rather than a much smaller random sample) of the seized illicit drugs. It is likely that a similar argument will now have to be considered by federal prosecutors if they are faced with presenting an estimate of  $Q$  to a jury using the beyond-a-reasonable-doubt standard. Experience so far has shown that prosecutors have not yet focussed on sample size as an important ingredient in the estimation process, and, in fact, sample sizes do not appear to have changed much since *Apprendi* was decided. The square-root rule is still king in many (if not most) forensic laboratories, and forensic scientists see no reason to change their current sampling practices even though the standard of proof has changed, potentially making convictions more difficult to achieve.

Prosecutors have also taken to provide a "summary witness" who will add up all the numbers for the jury. Jury verdict sheets, which previously only asked about the guilt of a defendant, now ask the jury, first, to determine guilt on the broadest charge, and then move on to more specific charges such as drug type and quantity, and even motivation. If defendants are convicted of multiple counts or crimes, prosecutors are asking judges to impose consecutive sentences rather than concurrent sentences, thereby reaching the greatest possible penalty regardless of enhancement.

### 6.3 *Recent Supreme Court Decisions Related to Apprendi*

The decision in *Apprendi* left open a number of questions, some related to how a defendant in a drug case is charged and what type of evidence regarding drug quantity is presented in court. Legal commentators and prosecutors are warning that *Apprendi* will lead to the

demise of the Federal Sentencing Guidelines. Furthermore, they say that recent Supreme Court decisions made post-*Apprendi* are threatening to make things even worse. Justice O'Connor also warned in her dissent that legislatures may try to sidestep the *Apprendi* ruling by rewriting their criminal statutes so that maximum penalties will be raised to cover all possible sentencing enhancements.

In the following three cases, questions raised by *Apprendi* were decided recently by the U.S. Supreme Court.

*Harris v. United States* (2002).

The question in *Harris* was whether there would be a constitutional problem if a judge, by a preponderance of the evidence, imposes a sentence which exceeds the applicable Federal Guidelines range, but is less than the applicable statutory maximum. Put another way, “if a judicial finding cannot be allowed to pierce the sentencing ceiling, can it logically be permitted to raise the sentencing floor, through the imposition of a mandatory minimum sentence?” (Greenhouse, 2002).

In this case, the defendant was convicted in 1999 for selling 4 ounces of marijuana to undercover officers in his pawnshop in Greensboro, North Carolina. The indictment charged him with drug dealing and carrying an unconcealed handgun, which would normally lead to a sentence of 5 years to life. The judge found that the defendant had actually ‘brandished’ the gun, which yielded a mandatory minimum sentence of 7 years. The Supreme Court in a 5–4 decision upheld a 4th Circuit Court of Appeals ruling that as long as the judge imposed a sentence that was within the statutory range, the defendant’s right to trial by jury was not compromised, even if the judge determined the facts that led to the higher minimum sentence.

*United States v. Cotton* (2002).

The major question in this drug case is whether *Apprendi* could be applied retroactively to other cases on collateral review. *Apprendi* was decided in the context of a direct appeal; the Supreme Court did not offer its opinion on whether *Apprendi* could be applied retroactively to cases on collateral review. Most federal circuit courts which have considered the issue decided that *Apprendi* is not retroactive to cases on collateral review unless the Supreme Court says so. Others have refused to consider the question of retroactivity until the Supreme Court weighs in with its opinion. In this case, the Supreme Court considered whether post-

*Apprendi* sentences which are dependent upon drug quantities uncharged in the indictment are subject to automatic reversal.

In this case, the trial judge sentenced five defendants to life imprisonment, while two others received sentences of 30 years. The 4th U.S. Circuit Court of Appeals, in Richmond, Virginia, found that the court did not have jurisdiction over an offense not charged in the indictment, vacated the sentences, and ordered all defendants to be resentenced to no more than 20 years. The Supreme Court, in a unanimous decision, reversed the appeals court, ruling that the appeals court had used the term ‘jurisdiction’ in a way that the Court felt was not what it means today (i.e., “the courts’ statutory or constitutional power to adjudicate the case”), and noted that a defective indictment does not deprive a court of its power to adjudicate a case. The Supreme Court went on to rule that the indictment’s omission of drug quantity was ‘plain error,’ that the error did not seriously affect the fairness, integrity, or public reputation of judicial proceedings, and that the evidence relating to drug quantity and conspiracy was ‘overwhelming’ and ‘essentially uncontroverted.’

*Ring v. Arizona* (2002).

The Supreme Court, in a 7–2 decision, overturned a ruling by the Arizona Supreme Court and decided that, post-*Apprendi*, at the sentencing/penalty phase of capital trials, juries (rather than judges) must make the determination that aggravating circumstances outweigh mitigating circumstances. The sentencing systems of nine states (Arizona, Colorado, Idaho, Montana, Nebraska, Alabama, Delaware, Florida, and Indiana) which allow a judge in a separate proceeding to find aggravating factors not considered by a jury that could increase a life sentence for a convicted defendant to a death sentence were declared unconstitutional. The Supreme Court’s decision means that only juries can turn a life sentence into an imposition of the death penalty.

## 7. ACKNOWLEDGEMENTS

The author thanks Dr Colin Aitken and Tom Nelson for information on Britain’s drug laws, and Dr Walter Rabl for information on Austria’s drug laws. This article was presented by the author as an invited lecture at the Fifth International Conference on Forensic Statistics held at Venice International University, Isola di San Servolo, Venice, Italy, from August 30 to September 2, 2002.

## 8. References

- Aitken, C.G.G. (1999). Sampling — how big a sample? *Journal of Forensic Sciences* 44, 750–760.
- Aitken, C.G.G. (2000). Sample size determination and evidence interpretation. In *Statistical Science in the Courtroom* (J. Gastwirth, ed.). Springer-Verlag, New York.
- Aitken, C.G.G., Bring, J., Leonard, T., and Papasouliotis, O. (1997), Estimation of quantities of drugs handled and the burden of proof, *Journal of the Royal Statistical Society*, 160, 333–350.
- Boswell, M.T., Gore, S.D., Lovison, G., and Patil, G.P. (1996). Annotated bibliography of composite sampling, part A: 1936–92. *Environmental and Ecological Statistics* 3, 1–50.
- British Standard BS 5309 (1993). Methods for the sampling of chemical products: Part I. Introduction and general principles.
- Bureau of Justice Assistance (1996). *National Assessment of Structured Sentencing*, NCJ-153853. U.S. Department of Justice, Washington, D.C.
- Cochran, W.G. (1977). *Sampling Techniques, Third Edition*. Wiley, New York.
- Colon, M., Rodriguez, G., and Diaz, R.O. (1993). Representative sampling of ‘street’ drug exhibits. *Journal of Forensic Sciences* 38 641–648.
- Desu, M.M. and Raghavarao, D. (1990). *Sample Size Methodology*. Academic Press, New York.
- Fienberg, S.E. (ed.)(1989), *The Evolving Role of Statistical Assessments as Evidence in the Courts*. Springer-Verlag, New York.
- Frank, R.S., Hinkley, S.W., and Hoffman, C.G. (1991). Representative sampling of drug seizures in multiple containers. *Journal of Forensic Sciences* 36, 350–357.
- Gastwirth, J.L. (1988), *Statistical Reasoning in Law and Public Policy: (Vol. 1) Statistical Concepts and Issues of Fairness; (Vol. 2) Tort Law, Evidence, and Health*. San Diego, CA: Academic Press.

- Gastwirth, J.L. (1992), Statistical reasoning in the legal setting. *American Statistician* 46, 55–69.
- Greenhouse, L. (2001), Court to weigh criminal sentencing, *New York Times*, December 11, 2001.
- Greenhouse, L. (2002), Justices weigh how far to take a sentencing revolution, *New York Times*, March 26, 2002.
- Greenwald, L. (1994), Relevant conduct and the impact of the preponderance standard of proof under the Federal Sentencing Guidelines: A denial of due process. *Vermont Law Review* 18, 529–563.
- Hedayat, S., Izenman, A.J., and Zhang, W.G. (1996), Random sampling for the forensic study of controlled substances (with discussion), *ASA Proceedings of the Section on Physical and Engineering Sciences*, 12–23.
- Izenman, A.J. (1991), Recent developments in nonparametric density estimation, *Journal of the American Statistical Association*, 86, 205–224.
- Izenman, A.J. (2000a), Introduction to two views of the *Shonubi* case, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed., 397–407. New York: Springer-Verlag.
- Izenman, A.J. (2000b), Assessing the statistical evidence in the *Shonubi* case, in *Statistical Science in the Courtroom*, J.L. Gastwirth, ed., 419–444. New York: Springer-Verlag.
- Izenman, A.J. (2001), Legal and statistical aspects of the forensic study of illicit drugs, *Statistical Science*, 16, 35–57.
- Izenman, A.J., Papasouliotis, O., Leonard, T., and Aitken, C.G.C. (1998), Bayesian predictive evaluation of measurement error with application to the assessment of illicit drug quantity, submitted for publication.
- Lancaster, V.A. and Keller-McNulty, S. (1998). A review of composite sampling methods. *Journal of the American Statistical Association* 93, 1216–1230.
- Särndal, C.-E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Tzidonoy, D. and Ravreby, M. (1992). A statistical approach to drug sampling: a case study. *Journal of Forensic Sciences* 37, 1541–1549.

United Nations (1987). *Recommended Methods for Testing Cannabis: Manual for Use by National Narcotics Laboratories*. United Nations, Division of Narcotic Drugs, Vienna.

United States Sentencing Commission (1995), *Federal Sentencing Guidelines Manual*. St. Paul, MN: West Publishing Company.

## 9. Cases Cited

*Apprendi v. New Jersey* (2000), 530 U.S. 466, 147 L.Ed.2d 435, 120 S.Ct. 2348, 2000 U.S. LEXIS 4304, 68 U.S.L.W. 4576.

*Harris v. United States* (2002), 153 L.Ed.2d 524, 122 S.Ct. 2406, 2002 U.S. LEXIS 4652, 70 U.S.L.W. 4655. See also 243 F.3d 806 (4th Cir., 2001) cert. granted, 122 S.Ct. 663 (2001) (No. 00-10666).

*in re Winship* (1970), 397 U.S. 358, 90 S.Ct. 1068, 25 L.Ed.2d 368.

*Patterson v. New York* (1977), 432 U.S. 197, 97 S.Ct. 2319, 53 L.Ed.2d 281.

*People v. Argro* (1975), 37 N.Y.2d 929, 379 N.Y.S.2d 840, 342 N.E.2d 601.

*People v. Ayala* (1981), 96 Ill.App.3d 785, 422 N.E.2d 127.

*People v. Black* (1994), 264 Ill.App.3d 875, 638 N.E.2d 233 (Ill.App. 1 Dist.).

*People v. Games* (1981), 94 Ill.App.3d 130, 418 N.E.2d 520.

*People v. Green* (1993), 256 Ill.App.3d 496, 628 N.E.2d 586 (Ill.App. 1 Dist.).

*People v. Hill* (1988), 169 Ill.App.3d 901, 524 N.E.2d 604, *appeal denied*, 122 Ill.2d 585, 530 N.E.2d 256.

*People v. Jones* (1995), 276 Ill.App.3d 926, 659 N.E.2d 70 (Ill.App. 1 Dist.), *aff'd*, 174 Ill.2d 427, 221 Ill.Dec. 192, 675 N.E.2d 99 (Ill. 1996).

*People v. Kaludis* (1986), 146 Ill.App.3d 888, 497 N.E.2d 360 (Ill.App. 1 Dist.), *appeal denied*, 113 Ill.2d 564.

*People v. Robinson* (1995), 167 Ill.2d 397, 212 Ill.Dec. 675, 657 N.E.2d 1020 (Ill.).

*R v Aramah* (1982), 4 Cr App R(S) 407.

*R v Aranguren* (1994), 16 Cr App R(S) 211.

*R v Morris* (2001), 1 Cr App R(S) 297.

*R v Warren and Beeley* (1996), 1 Cr App Rep 120.

*Ring v. Arizona* (2002), 153 L.Ed.2d 556, 122 S.Ct. 2428, 2002 U.S. LEXIS 4651, 70 U.S.L.W. 4666.

*State of Minnesota v. Robinson* (1994), 517 N.W.2d 336 (Minn.).

*State of North Carolina v. Hayes* (1976), 291 N.C. 293, 230 S.E.2d 146 (N.C.).

*United States v. Cotton* (2002), 152 L.Ed.2d 860, 122 S.Ct. 1781, 2002 U.S. LEXIS 3565, 70 U.S.L.W. 4429.

*United States v. Fatico* (1978), 458 F.Supp. 388 (E.D.N.Y.), *aff'd on other grounds*, 603 F.2d 1053 (2d Cir. 1979), *cert. denied*, 444 U.S. 1073, 100 S.Ct. 1018, 62 L.Ed.2d 755 (1980).

*United States v. Gigante* (1996), 94 F.3d 53 (2d Cir.).

*United States v. Lam Kwong-Wah* (1992), 966 F.2d 682 (D.C. Cir.), *cert. denied*, 113 S.Ct. 287, 121 L.Ed.2d 213 (1992).

*United States v. Royal* (1992), 972 F.2d 643 (5th Cir.).

*United States v. Shonubi* (1997), 962 F.Supp. 370 (E.D.N.Y.). See also 103 F.3d 1085 (2d Cir. 1997), 895 F.Supp. 460 (E.D.N.Y. 1995), 998 F.2d 84 (2d Cir. 1993), 802 F.Supp. 859 (E.D.N.Y. 1992).