Abstract

Data on illicit drug purity and prices are invaluable but problematic. Purists argue they are unsuitable for economic analysis (Manski et al., 2001; Horowitz, 2001), but in reality they are used frequently (ONDCP 2001a, 2001b, 2004; Grossman, 2004). This paper reviews data and conceptual issues that people producing, analyzing, and consuming drug price and purity series should understand in order to reduce the likelihood of misinterpretation. It also identifies aspects of drug markets that are both poorly understood and relevant to some of these issues. They constitute a useful research agenda for health and law enforcement communities who would benefit from better data on the supply, availability, and use of illicit drugs.

1.0 Introduction

Illicit drugs are, ultimately, consumer goods that are produced and distributed through markets, so it is natural to want to examine price, purity, and supply in those market (Caulkins and Reuter, 1998). Data on price, purity, and supply are complicated and problematic. Indeed, some argue that the best-known and most widely used source of data on drug prices, the Drug Enforcement Administration’s (DEA’s) System to Retrieve Information from Drug Evidence (STRIDE) should not be used for economic analysis (Manski et al., 2001; Horowitz, 2001).

The price and purity series derived from such imperfect sources do, however, strongly correlate with independent series that are seen as relevant indicators of drug-
related phenomena, specifically series based on emergency department mentions, medical examiner mentions, surveys of the household population and high-school seniors, and urinalysis results for both arrestees and the workforce generally (Hyatt and Rhodes, 1995; Crane et al., 1997; Saffer and Chaloupka, 1999; Caulkins, 1999, 2001; DeSimone, 2001; DeSimone and Farrelly, 2003). Hence, it seems that with proper methods one can extract some information from these data sources, and as any good Bayesian would assert, it is typically better to have some information rather than none.

The goal of this paper is to review conceptual issues relevant to responsibly interpreting data on illicit drug price and purity. The next section begins with a brief review of three broad types of applications of price and purity data. The subsequent and largest portion of the paper is devoted to unusual characteristics of markets that bear on the creation and interpretation of drug price and purity series. The fourth section distinguishes three types of sources of price and purity data, and notes some of their relative advantages and disadvantages.

2.0 Uses of Price Data

Caulkins and Reuter (1998) review the multiple uses of drug price and purity data in detail, but to simplify, they are used in three types of research projects: (1) correlating changes in measures of use with changes in price to estimate price-responsiveness, e.g., the price-elasticity of demand, (2) correlating changes in supply indicators (purity, price or price markups) with changes in measures of enforcement to determine enforcement’s effectiveness at disrupting markets, and (3) using the absolute level of price and/or purity as a “conversion factor” to reconcile estimates of disparate quantities.

The first (regressing use on price, broadly construed) has developed into quite a large literature in economics (recently reviewed by Grossman, 2004; see also Chaloupka and Pacula, 2000). The methods employed have advanced considerably from simple correlations a decade ago to sophisticated econometric models, but there are still limitations. Notably, the most papers are written for the drugs and populations that are associated with the least severe social problems, primarily because the data are better for, say, marijuana use among high school seniors, then they are for cocaine and heroin use by criminally-involved dependent users.
The principal conceptual contribution of this literature has been to establish beyond doubt that drug use does indeed respond to variation in price even for dependent users. At a pragmatic level, the various elasticity estimates are crucial input parameters to “systems models” of drug policy effectiveness. Those quantitative models would benefit from (1) more precise estimates, (2) divisions of the overall elasticity into short-term elasticity of per capita consumption, elasticity of initiation, elasticity of desistance, and, in some cases, elasticity of escalation, (3) cross- as well as own-price elasticity estimates, and (4) better estimates for heavier use of harder substances by more troubled populations. Fulfilling those needs will provide fodder for quite a few more papers in this already substantial literature.

The second type of application (loosely, regressing price changes on changes in enforcement) has engendered a smaller but still important literature (DiNardo, 1993; Crane et al., 1997; Yuan and Caulkins, 1998; Kuziemko and Levitt, 2001; Bushway et al., mimeo; Pietzmann, forthcoming). The recent flurry of papers on the Australian heroin drought (e.g., Weatherburn, et al., 2002; Dietz and Fitzgerald, 2002; Day et al., 2003; Smithson et al., 2004) represent something of a special case inasmuch as they evaluate a supply-side disruption’s impact on market indicators, but it is not yet clear what caused the supply disruption. A principal limitation of this literature is that many price estimates are annual or at best quarterly, whereas markets adapt so quickly to all but the largest disruptions that higher-frequency price series would be more useful. Likewise, price and purity series are not always available at the spatial unit of aggregation that maps neatly into the footprint of the supply disruption.

The third type of application of drug prices is the least well-defined and is perhaps best described by example. For various reasons it is important to develop demand-side as well as supply-side estimates of the total volume of drugs produced and consumed (Drug Availability Steering Committee, 2002). The Office of National Drug Control Policy’s (2001) standard method for demand-side estimation starts with total spending on the drug (derived as the product of the number of users times spending per user) and divides by the average retail price to translate the spending estimate (in dollars) to a consumption estimate (in kilograms). Another classic example is estimating the proportion of retail
drug sales revenue that remains in the US vs. going to source countries by comparing retail to import prices.

Some papers (e.g., Caulkins and Reuter, 1998) sit at the intersection of the second and third categories inasmuch as they try to explain price mark-ups along the distribution chain in terms of various production costs, including compensation for the risks of enforcement (cf., Reuter and Kleiman, 1986).

The most salient issues concerning the quality of price and purity data vary by type of application. To give one simple example, if price estimates are systematically 20% too high, that matters enormously to efforts to estimate quantities consumed from spending estimates but not so much in regression analyses that focus on price changes rather than levels. Indeed, if all price estimates were inflated by exactly 20%, that would have no impact whatsoever on resulting estimates of elasticities of demand, which relate percentage change in consumption to percentage change in price. Conversely, fluctuations in enforcement priorities could generate spurious blips in price series measured with high frequency (e.g., monthly), but if those fluctuations averaged out over the course of a year, estimates of absolute levels of prices for the year overall (the typical input in the third type of application) might not be affected much at all.

Rather than proceeding directly to lists of these individual observations, it is more constructive to first review some basic characteristics of drug markets that distinguish them from more traditional markets.

3.0 Special Characteristics of Drug Markets

Drug markets are unmistakably markets. They are populated by buyers and sellers. Prices fluctuate in response to market pressures and balance the quantity supplied with the quantity demanded, at least in the medium- and long-run.

Nevertheless drug markets differ from more familiar markets in ways that are essential to understand if price and purity data are to be interpreted properly. A great deal has been written about the characteristics of particular drug markets (Natarajan and Hough, 2000), particularly for low-level market from an ethnographic perspective (e.g., Bourgois, 1996, 1998; Curtis and Wendel, 2000; Johnson et al., 1995, 2000; Fagan, 1992; Furst et al., 1999, 2004; Mieczkowski, 1992, 1994). What is more relevant for present
purposes, however, is reflections on general characteristics of typical drug markets at multiple market levels from an economic or a policy analytic perspective. What follows is an abbreviated set of some of the essential differences, drawn to an important degree from an excellent parallel discussion in Kleiman (1992) and from various writings of Peter Reuter.

3.1 Conventional Pricing and Non-Conventional Contents

For many purposes one is particularly interested in retail price variation. However, for over four decades conventional pricing has characterized illicit markets (Wendell and Curtis, 2000), so superficially there is no such variation, with almost every retail transaction being for a standard price. The amount of heroin in a “dime bag” might change over time, but by definition a dime bag always costs $10 (“dime” being street argot for $10).

“Dime bags” and “nickel rocks” are extreme examples, but even when the price is defined by the name of the transaction size, there is still a strong tendency for retail transactions to predominantly be for certain, standard, round amounts. For example, in 2000, three-quarters (80/106) IDRS respondents in Victoria, Australia who reported the price of their last “rock” or “cap” purchase cited a value of $50, and just four of the 106 figures cited were not round multiples of $5.¹

At the same time, what one obtains in a drug purchase is not standardized, either with respect to quantity or quality. One typically buys milk in standardized quantities (pint, quart, half-gallon, or gallon), and the quantity in the container will essentially always be very, very close to the advertised quantity. Quite literally, if one drew a frequency distribution of transaction sizes for milk, there would be four bars at those four quantities and nothing in between.

The picture looks very different for illicit drugs. (See Arkes et al., 2004, for examples.) The frequency distribution shows spikes around certain semi-standard quantities, such as 1 gram, “eight-balls” (one-eighth ounce), one ounce, and one-eighth kilogram for cocaine. However, there are many more than just four spikes. For example, powder cocaine also has discernable spikes at ¼ ounce, ½ ounce, 2 ounces, quarter kilo,

¹ Author’s analysis.
half kilo, and one kilo. Furthermore, the spikes are not as sharp as they would be for milk. “One ounce” purchases range from about 23 – 30 grams with the modal quantity being a bit below 28.35 grams. (The discrepancies between claimed and actual weight are not normally distributed with zero mean; there is a clear bias toward selling on the light side of a round amount.) Finally, in between the round sizes there is an almost continuous non-zero smear on the frequency diagram. One essentially never observes a retail transaction for an odd quantity such as 1.6 pints of milk, but there are 1.6 gram and 1.6 ounce cocaine transactions.

The variation from package to package is by no means confined to total weight. There are also quality differences, as is discussed further below.

The implication of conventional pricing coupled with non-standardized transaction contents is that the price per unit adjusts not primarily by having the nominal price change, but rather by variation in the amount of drugs obtained for that standard dollar figure. When prices change dramatically the standard dollar figure may vary; “nickel” rocks replaced “$20 rocks” as cocaine prices fell during the 1980s. But in order to develop price series with any degree of precision, one has to divide the sticker price by the amount obtained in order to obtain some sort of normalized price per unit of drug purchased.

3.2 Quantity Discounts and Quality Premia

The most important and most discussed set of issues pertaining to estimating drug price series pertain to quantity discounts and quality premia (Caulkins and Padman, 1993; Caulkins, 1994; Clements, forthcoming). These issues are reviewed in this subsection. Readers familiar with these ideas should skip to the next subsection.

3.2.1 Quantity Discounts

For a wide range of illicit drugs, price per unit increases markedly as the drugs move down the distribution chain or, looking at from the customers’ perspective, substantial quantity discounts are the norm. Specifically, price per unit is proportional to transaction size raised to a (negative) power (Caulkins and Padman, 1993; Clements, forthcoming).
Quantity discounts are observed for many goods, but the discounts are far more substantial for illicit drugs. One implication is that when trying to make dollar costs and flows “add up” for illicit markets, it is essential to value seizures at the value per unit appropriate for the market level in question; there can be a significant difference between the replacement cost of drugs and their retail value. (Reuter, 1988)

Another implication is that one cannot estimate the dollar value of the retail market simply by multiplying quantity consumed by the price per gram; the distribution of retail purchase sizes must be factored in as well. Further, if that purchase size distribution varies over time, using the price per gram times the quantity consumed can distort trends in the estimated dollar value of the market, as well as the absolute value. For cocaine, such distortion could have been on the order of 40% between the early and late 1980s. (Caulkins, 1994b)

3.2.2 Illicit Drugs Are of Varying Quality

Essentially all kinds of drugs exhibit considerable variation in quality. Powder drugs such as cocaine and white and brown heroin have traditionally between both diluted and adulterated as they moved down the distribution chain (Moore, 1973). At one time this was done to a considerable extent. For example, retail heroin purity in the US in the 1970s was in the single digit percents, even though wholesale purity might well have been on the order of 50%. Such dilution and adulteration is less pronounced than in the past (ONDCP, 2004a), and Coomber (1999) argues that much of the still non-trivial variation in purity stems from variation introduced at the production stage.

Methamphetamine purity in the US has varied dramatically over the last 15 years with multiple peaks and valleys (Caulkins et al., 2004), very possibly driven at least in part by successive waves of precursor regulations (Cunningham and Liu, 2003).

There are many reports of MDMA being adulterated (Hanson, 2002). Indeed, some maintain that a significant number of the adverse reactions experienced by people who believed they were taking MDMA were in fact caused by adulterants. (NEED CITES)

Marijuana is rarely diluted or adulterated, but its potency, at least as reflected by THC content, varies considerably, both from sample to sample and over time (ElSohly et
There are also systematic quality differences between types of cannabis products, e.g., “sinsemilla” vs. “commercial grade” vs. “ditch weed”, that are reflected in quoted prices (Kleiman, 1989).

The upshot of this is that one should seek to examine quality-adjusted prices, not simple or “raw” prices. Since price changes often manifest through quality changes, particularly for cocaine, heroin, and methamphetamine, failing to adjust for quality variation tends to understate variation in the effective price of drugs, often dramatically.

To illustrate, the Australian heroin drought increased the price per gram of heroin in Victoria unadjusted for purity from $300 per gram in 2000 to $400 per gram in March/April 2001 (Miller et al., 2001) and $500 per gram in June/July 2001 (Fry et al., 2002). At the same time, the average purity of heroin seized in Victoria dropped from 46.7% in 2000 to 26.0% in March/April 2001 and 14.9% by June/July. If one ignored quality variation, the price increases were 33% and 67%. Factoring in the change in quality, the price increases were 139% and 422% -- a quintupling of the price per unit of intoxicant.

3.2.3 Illicit Drugs Are a “Double” Experience Good

Economists use the term “experience good” to describe something whose quality is not known to the buyer until after the price has been negotiated and the purchase has been consummated. That classic example is a restaurant meal. Certain things about the meal are known from the menu at the time it is ordered, but not everything about how it will taste. As a result, sometimes customers end up disappointed, wishing that they had not paid so much or even patronized that restaurant at all.

Likewise, drug buyers do not know at the time of purchase what is the drug’s quality (primarily purity, but to a lesser extent other attributes as well). That is not surprising at the retail level. Transactions are rushed; formal testing is time consuming and may be impractical for small quantities.

What is somewhat surprising is that apparently it seems also not to be common to quantitatively assay purity for wholesale purchases either (Reuter and Caulkins, forthcoming), even though those transactions involve higher stakes and are more likely to take place in a hotel room or other discrete site.
An obvious consequence of wholesale buyers not knowing the quality of what they purchase is that not only the retail purchasers but also the retail sellers are at best imperfectly informed as to quality, since retail sellers are themselves buyers at the next higher market level. (The retail sellers would still have somewhat more information than their customers because they would know whether they themselves “cut” the drugs, even if they did not know what was the original quality of the drugs they obtained.)

For drugs such as cocaine and heroin there are not just two market levels made up of retailers and wholesalers. Typically cocaine undergoes something like 5 transactions between its import in multi-kilo to multi-hundred kilo lots and its retail sale in units of 0.1 gram to a few grams. Hence, not just users and retail sellers, but also lower-level wholesale sellers may have quite imperfect information about purity. In this sense, drugs are a “double” experience good.

The fact that drugs are an experience good combined with the substantial variation in quality observed in drug markets has a very important implication for the analysis of transaction-level price data. In particular, it is not safe to assume that the customer knowingly and willingly paid the observed amount for the observed pure quantity of drugs obtained. Simply put, drug purchasers can be just as disappointed and regretful as a restaurant patron after a bad meal. Indeed, given the emphasis on quality in legitimate markets over the last few decades and the lack of legal recourse for customers swindled in illegal markets, drug purchasers are regretful far more often. The coefficient of variation on variation in drug quality is truly enormous compared to that in legal markets (Reuter and Caulkins, forthcoming).

To make this point concrete, if one observes that someone paid $100 for 1 gram of cocaine that was 10% pure, that does not mean that the going price per pure gram is $100 / (1 * 10%) = $1,000 per pure gram. If the usual purity for transactions of that size at that time, place, etc. were 67%, the customer paying the $100 probably expected that gram of powder to be about 67% pure cocaine and it would be more accurate to say that

---

2 Furthermore, one or more brokers and several different smugglers likely handle the cocaine between the lab where it is produced and its arrival in the US. They are perhaps less likely to cut the drugs since most do not actually own the drugs they are transporting and high potency facilitates concealment, but it is at least theoretically a possibility.
the customer’s understanding of the market price were $100 / (1 * 67%) = $150 per pure gram.

This concept (called the “Expected Purity Hypothesis”) and its implications for creating price series are discussed at length in Caulkins (1994) and are routinely incorporated into generation of price series.

Failing to recognize that drugs are experience goods can distort estimates of price trends. In the past it was not uncommon to simply regress price paid on transaction size and purity, or, more typically, to do so with logged quantities (Rhodes et al., 1993). The resulting regression coefficients for purity were typically rather small (Caulkins and Padman, 1993) because the quality surprises amount to measurement error, which biases the purity coefficient toward zero. This bias often makes purity’s regression coefficient much smaller than the coefficient on quantity. Such substantial inequality should make one suspicious; it suggests that all drug sellers have easy opportunities to increase profits by further diluting their drugs, essentially using their own stocks to arbitrage the value difference between low and high purity markets.

If these regression models were applied to data series whose purity varies over time, the effect of that change on the true price per pure gram would be understated. In particular, if purity rose over time (as it did dramatically in the 1980s for cocaine and through the early 1990s for heroin), then the estimate time series for price per pure gram would not fall as much as the actual price per pure gram did. This effect is illustrated in the contrast between the baseline and “alternate” price series produced by ONDCP (2001) a few years ago.

3.3 Spatial Variation in Market Prices

It has long been recognized that there is substantial variation in drug prices not only as one moves down the distribution chain but also between locations even at similar market levels across countries (UNODC, 2004), within the US between cities (Caulkins, 1995), between different neighborhoods within the same city (Weatherburn and Lind, 1997; Caulkins and Reuter, 1998), and even at the wholesale level (Caulkins, 1994). Given drugs’ very high value per unit weight, one might initially expect arbitrage to eliminate quickly such spatial variation. Apparently, however, information flows are
sufficiently imperfect and/or lateral transactions costs sufficiently great that these differences are not closed. Indeed, Reuter describes the case of a dealer who made handsome profits almost entirely by exploiting such spatial variation in price, selling essentially the same quantities as he was buying (Reuter and Haga, 1989).

Furthermore, prices in different cities in the US are far from being perfectly correlated, with prices rising in one city at the same time they are falling somewhere else. This imperfect correlation is less pronounced for powder cocaine, which may have more of a national market, than it is for crack or heroin (Caulkins et al., 2004). Methamphetamine may be the major drug whose markets display the greatest regional variation in the US (Caulkins, 2003).

On the one hand, this spatial variation can be a blessing to researchers seeking to establish correlations between prices and various drug use-related outcomes. Indeed, given the number of unmeasured time-varying, confounding correlates, a panel-data approach to estimating the elasticity of demand is very strongly preferred over a simple time series based approach.

On the other hand, this spatial variation greatly complicates the seemingly straightforward task of answering a question as simple as, “What is the retail price of cocaine (or some other drug) in the US?” To begin with, it is clear that one should not refer to a single price, but rather to an average. But exactly what average should be taken? One could, for example, take the weighted average of the price in 25 different cities, weighting by each city’s population, but is that the “right” weighted average?

Three distinguishable issues complicate this question of averaging. The first is that one does not necessarily have good city-level price data for all locations one might wish to include in the average. If one ranks cities based on the number of price observations they have in STRIDE, the number of observations per city drops off rather quickly once one moves beyond the first ten or twenty cities, dramatically so if one wants to estimate prices with greater time resolution than simply by year. Hence, in computing a national average price, one is often faced with a no-win trade-off between using a small number of cities which do not fully represent the US as a whole or using a larger number of cities, by including ones for which one may not have enough data points to overcome the great variability in price and purity.
The second issue is conceptual and subtle. Even if data were perfect, there is not just one “right” weighted average. One of the more common uses of a national retail price estimate is converting from total spending on drugs to a “demand-side” estimate of total consumption. Divide national retail spending by national retail price, and you get national consumption. For that purpose one would like the weights in the weighted average to reflect relative quantities consumed in each city. If twice as much heroin was consumed in Washington DC as in Pittsburgh, then Washington’s retail price ought to receive twice as much weight as Pittsburgh’s retail price in computing the weighted average national price.

That same weighting is not unreasonable and has in the past been used for producing general-purpose national price series. However, the latest ONDCP report takes the simpler and more transparent approach of weighting by city population. The resulting series might be interpreted as the national average price observed by potential users (living in cities with sufficient data to generate city-specific prices).

Suppose, however, one wanted to explore policy’s success at protecting youth from the temptation to try drugs. Then it would make more sense to weight each city’s retail price by the number of people in that city who are susceptible to initiating the drug, perhaps proxied by the number of 15-25 year olds in the city. Likewise, if the focus were on prices and incentives to enter treatment, the appropriate weighting might be by the number of dependent users in each city.

The third issue is that whatever weights one would theoretically like to use, there almost certainly do not exist data on those weights, so one is forced to use a proxy, often a rather poor one. For example, Abt used DAWN to develop a proxy for the amount of use by city when computing its weighted average national price (ONDCP, 2001). This presents a range of problems including that DAWN cachement areas not always matching the areas in STRIDE associated with a particular city and the simple fact that DAWN data are not collected for all cities, so Abt was forced to proxy DAWN mentions in one city with those in a nearby city (e.g., using Seattle data for Portland, adjusting for population).\(^3\)

---

\(^3\) Some of the resulting city-specific weights looked suspicious. For example, based on DAWN, in 2002 cocaine use in St. Louis would seem to be greater than in Washington DC, whereas state-level estimates from the 2001 NHSDA suggest that this is unlikely.
Just to add one further complication, there can be systematic variation across cities in the size of a typical retail transaction which, coupled with quantity discounts, means that ratios of amounts spent to amounts purchased may be city-specific. For example, heroin in New York is customarily sold in $10 “dime bags” whereas in San Francisco it is more often sold in $20 bags (Agar et al., 1998).

3.4 Variability as a Predictor Independent of the Mean

As discussed above, prices vary over time, across locations, and between market levels. Purity does as well. Distinct from these is the variability from transaction to transaction even controlling for time, location, and market level. That residual variability is dramatically larger than for conventional goods and is not necessarily uniform across drugs or time (Reuter and Caulkins, forthcoming).

One of this enormous variability is that variability in price and/or purity might itself be an independent predictor of use-related outcomes and component of the dependent variable in analysis of supply disruptions.

A concrete example that has been much discussed but, to the best of the author’s knowledge, never thoroughly tested empirically is the role variability in purity plays in driving overdoses. One hypothesis is that overdoses increase when the (mean) price per pure gram falls because lower prices lead to greater use. A second hypothesis is that increasing (mean) purity leads to more overdoses. A third and conceptually quite distinct hypothesis is that greater variability in purity predicts overdoses, even with no change in the mean price or purity, because users inadvertently consume more pure drug than they intend to when they happen to purchase drugs that are more pure than is typical of that time, location, and market level. Indeed, one might guess that it is not literally the standard deviation or some other general purpose measure of variability that matters, but rather the ratio of the $97^{th}$ or $99^{th}$ percentile of the purity distribution to the average that might predict numbers of overdoses.

A second largely unexplored implication of the enormous variability in price per pure gram is that purchase risk may become important to the utility customers derive from patronizing the market. Because of the variability, the effective cost of drugs is higher for individuals who are risk-averse than it is for those who are risk neutral. For
example, suppose the actual purity distribution is uniform between 40% and 80%. A risk-averse person would be indifferent between that situation and purity being consistently 60%, but for someone who is risk-averse, they might be willing to accept a lower average purity (e.g., 55%) if they could avoid the risk associated with such variability. Inasmuch as most people are risk-averse to at least some extent, increasing variability might be indistinguishable from a reduction in average purity in terms of the value to the customer and, hence, the incentive to reduce consumption.

3.5 Non-Monetary Costs of Obtaining Drugs

The dollar price does not reflect all of the economic costs of acquiring drugs, even after adjusting for quantity and quality. The most discussed non-monetary costs are the “search time” or “shoe leather” costs of locating a supplier and consummating a transaction. This concept was pioneered by Moore (1973) who pointed out the policy benefits of increasing such non-dollar retail costs relative to driving up the dollar costs. Kleiman (1988) further developed these ideas, and with Rocheleau and Boyum (1994) showed that data on search time could be collected systematically.

Just as a complete model of price might best include measures of variability, so too it ought to reflect some measure of search time costs, whether the price vector were on the right-hand side of a regression predicting measures related to use or on the left-hand side of a regression whose predictors are enforcement or other measures of supply disruption. Practically speaking, this is possible when the price data derive from extended interviews with users, but not when the price data come from systems such as STRIDE or synthetic judgments of law enforcement officials.

An entirely distinct aspect of non-monetary costs is payments in kind. Bartering goods or services for drugs is commonly reported in the ethnographic literature, in at least three forms: trading sex for drugs, drug sellers who also operate as fences accepting stolen goods as payment, and drug distribution workers being paid in drugs for their services (both “jugglers” who sell on consignment and low-level workouts such as touts and look-outs).

Payments in kind pose a challenge for efforts to estimate quantities consumed from amounts spent on drugs and prices. In particular, if estimates of the amounts spent...
reflect only dollars spent, and do not include the monetary value of goods and services traded for drugs, then the resulting demand-side consumption estimates will be biased downward. This potential bias is recognized and crude adjustments made for it (ONDCP, 2001), but there is as of yet still only imperfect information concerning how widespread this bartering is for the nation overall.

The issue becomes further complicated when the seller and customer are connected in ways beyond an arm’s length business relationship. How does one think about the “price” when a dealer or even a user provides drugs for free to a girlfriend?

For marijuana this is a central, not a peripheral issue. The 2001 National Household Survey on Drug Abuse asked past-year marijuana users about their most recent marijuana acquisition. Fully 87% reported last acquiring marijuana from a friend or relative and 58% said they got it for free (Caulkins and Pacula, in submission). So for most marijuana users, the “price” of marijuana is essentially invisible, and is probably effectively paid in reciprocal acts of generosity embedded within a social relationship of which marijuana distribution is only a small part. Someone might give a friend rides to work; in return, that person might supply the driver with free marijuana on weekends. In such a context, how does one think about the retail price? Perhaps it is still legitimate to estimate the effective price for all users based on the minority who pay exclusively in dollars. If the dollar price of marijuana rises, perhaps the friendly supplier in the example above would implicitly demand more car rides per gram given? At a minimum, however, it is important to recognize that, particularly for some drugs, many retail transactions may be embedded in a social relationship rather than being arm’s length business transactions.

4.0 Sources of Price Data

There are at least three types of data concerning price and purity, that offer some distinct advantages and disadvantages. Two of the three are reviewed briefly, then the bulk of the paper addresses issues with the third and most commonly used in academic research.

4.1 Synthesis Reports by Government Agencies:
Some of the earliest price data in the US at least came from annual reports from government agencies (typically law enforcement), such as the “NNICC reports” from the National Narcotics Intelligence Coordinating Committee. Other examples include reports from the Regional Information Sharing System (RISS; http://www.iir.com/riss/) including the Western States Information Network (WSIN) and the Middle Atlantic-Great Lakes Organized Crime Law Enforcement Network (MAGLOCLEN) and from the United Nations Office on Drugs and Crime (UNODC, 2003, 2004). An example from overseas is the Australian Bureau of Criminal Intelligence (ABCI) Australian Illicit Drug Report (ABCI, 2002).

Most of these documents describe the price and purity for various drugs and transaction sizes, sometimes for specific quantities (e.g., one kilogram), sometimes in “street” quantities (e.g., an “eight-ball”), sometimes just for broad market levels (e.g., “retail” and “wholesale”). Some give point estimates. Some report ranges. Some a mixture of the two, not always with an apparent reason why some but not other estimates include ranges. Rarely do the reports include even a rudimentary explanation of the methodology underlying these estimates. Often, as in the case of UNODC reports, they can represent a synthesis of information passed up from other reporting units. Anecdotally, in at least some instances the ranges are literal; that is, they represent the range from the lowest to the highest figure observed, even if the number of “data points” is large and some sort of trimmed range or inter-quartile range might be more informative and less affected by outliers.

Besides the lack of information concerning the underlying methods, an additional concern is that the methods might change over time without any documentation concerning those changes. For the US, given the number of reorganizations before 1988 affecting of which agency was the lead federal drug coordinating organization (equivalent to the current Office of National Drug Control Policy or ONDCP), this is particularly a concern for older documents, but ironically, it is precisely in those historical periods that other types of data are also weakest.

These data can certainly be used for statistical analysis (e.g., Caulkins and Padman, 1993; Caulkins, 1995; Farrell et al., 1996).
4.2 Self-Report by Users

Various sources describe self-reported price data, ranging from the US National Survey on Drug Use and Health (NSDUH) questions about marijuana markets (e.g., Caulkins and Pacula, in submission), to the Australian Illicit Drug Reporting System’s interviews of needle and syringe exchange patrons (Miller et al., 20001; Fry et al., 2002), to exotic sources such as the “High Times” magazine’s index of cannabis prices.\(^4\)

Self-report data raise several issues. One of the most important is that users generally have no way to quantify the purity of what they purchase (cf., the Australian heroin drought purchase price estimates above). Indeed, users have somewhat imperfect ability to describe the quantity obtained. As noted, a “street gram” is, on average, a little light of a gram.

There are also, of course, all of the usual issues surrounding self-report, such as the potential for users to distort their answers. It is possible, for instance, that under-reporting might be correlated with prices paid, e.g., if heavier users are more likely to lie about their use and are more likely to know where to get better deals. There could also be less strategic distortions; people who want to create the impression that they are street-saavy shoppers might report with relish a great deal obtained last week while glossing over their getting burned the previous day because they were in such a hurry to score. There has been a great deal of research on the validity of self-reported drug use (e.g., Harrison, 1995), but less on the validity of self-reported price data.

A distinct and fundamentally important issue pertains to what exactly is being sampled by most self-report data. The household survey gives fantastic information concerning the last purchase of a random sample of past-year marijuana users, but next to nothing about typical past-year purchases. To make the distinction clear by stylized example, suppose that half of marijuana users make one purchase per year at a price of $300 per gram and half purchase daily at a price of $100 per gram. The average of the transaction reported in a household survey of this population would be \((300 + 100) / 2 = 200\) but the average price of a transaction would be \((1/year * 300 + 365/year * 100) / (366) = 100.55\).

\(^4\) Rosalie Pacula collected High Times data over a number of years, but the author is not aware of their yet being used in academic research.
One cannot simply adjust for this by weighting respondents’ answers by how often they purchased in the last year because of the problem of random incidence (Larson and Odoni, 1981). Larger and more expensive purchases might more often be followed by longer than average inter-purchase times and, hence, be more likely to be the most recent purchase made at the time of the survey. Conversely, small purchases followed by shorter inter-purchase times are less likely to be sampled by a survey whose timing is independent of the purchasing processes.

4.3 Transaction-Level Data

Law enforcement agencies sample illicit drug market prices routinely in the course of their operations, most notably through “buy-bust” operations. Dollar amounts spent on these transactions are typically recorded carefully, if for no other reason than to minimize the risk of corruption concerning misuse of “PE/PI” (purchase evidence/purchase information) money. Weights are also recorded fairly carefully given the number of sentencing laws that base punishments on the quantity possessed, and in certain cases law enforcement agents performance reviews are based on the “quality” of cases made, of which drug weight may be a contributing determinant.

For samples obtained by many local police departments there is no quantitative assay of purity or potency. A simple chemical test is performed to confirm the presence of an illicit substance, but since sentencing statures are typically based on the total weight of the material containing an illicit drug, not on the pure quantity of the drug present (Caulkins et al., 1997), there is often on need for a more careful analysis.

However, when the samples are sent to a criminal forensics lab, a wealth of other information may be obtained including not just the purity of the principal drug present, but also the presence and amount of various diluents and adulterants and sometimes information about the “source” of the substance, e.g. for heroin whether its chemical “signature” suggest that it was produced by a chemist using the methods typical of Southwest Asia, Southeast Asian, Mexico, or South America.

The most commonly used source of price data for recent analyses is of this type. It is the Drug Enforcement Administration’s System to Retrieve Information from Drug Evidence or STRIDE database (Frank, 1987). STRIDE has both great strengths, notably
its large numbers of transaction-level data from across the country that have been recorded in a more or less consistent manner over many years. (There are inconsistencies over time, for instance, in the geographic coding, but those issues can be dealt with.) STRIDE’s principal disadvantage is that it is not collected for research purposes. It is an administrative data set whose sampling reflects law enforcement activity and prerogatives. It does not contain a representative, let alone random, sample of the drug transactions. Manski et al. (2001) and Horowitz (2001) have been sharply critical of STRIDE for this reason, but it may still be the best available source of data.

From the perspective of a scientific “purist”, STRIDE might be fatally flawed, as would be almost any data source on illicit drug prices. However, for those who believe that some information is better than none, STRIDE can be useful, particularly if its ideosyncracies are understood and appropriate methods are employed (Caulkins, 2001b). In some sense STRIDE is not a particularly complicated data set, but it does not have a codebook, so its responsible use requires acquiring a certain degree of tacit knowledge.

To give just two examples, STRIDE contains a “Form” field that specifies the measurement units for a sample (grams, ounces, etc.). However, not all past releases of STRIDE have included this field and it has not been uncommon to implicitly assume that all observations are measured in units. For cocaine this has been a minor issue since only about one percent of cocaine observations are measured in something other than grams, but 16 percent of methamphetamine observations are, and earlier price-series for methamphetamines did not always correlate as well with exogenous data as do more recent series (cf., ONDCP 2001, 2004).

As a second example, STRID has sometimes been distributed as a semi-colon delimited file, but at least two STRIDE fields do occasionally contain semi-colons for some records. Colon-delimited downloads appear to be more reliable.

It is not of interest here to delve into such minutiae. For that, the reader is referred to ONDCP (2004) as well as various papers in the economics literature on estimating price elasticity of demand. There are, however, a few conceptual issues concerning STRIDE or STRIDE-like data sets that bear mention.

The first is simply to realize that the typical STRIDE observation stems from law enforcement making a purchase in order to arrest the seller. It is common, therefore, to
worry that drug agents do not know much about the true market price or not to bargain as hard as a real market participant. On the other hand, the agent has to behave enough like a real market participant to trick the seller into believing the agent is a real market participant. No seller would say, “Ah. This guy is offering to pay so much he must be a drug agent. Never mind that I’ll get arrested if we complete the transaction. I can’t resist the chance to make more than the usual profit on this sale.”

On the other hand, actual drug market participants may have regular “connections” (Moore, 1973) with whom they enjoy an ongoing “bilateral monopoly” (Kleiman, 1992), in part as a way of solving the problem of variability in quality by making the inherent “Prisoner’s Dilemma” a repeated game (Reuter and Caulkins, forthcoming). In contrast, undercover agents would typically only purchase from a given seller once or at most a few times before making an arrest. If first-time buyers are routinely charged a price premium, then prices from STRIDE or any other law-enforcement data base might be systematically biased upward.

A second issue is the absence of potentially relevant covariates. DEA case-files record a wealth of information concerning, for example, the time of day of the purchase; whether the purchase was indoors, in a vehicle, or outside; the demographic characteristics of the seller; and the length and nature of the relationship between buyer and seller. Such information is not included in STRIDE, however, and coding it from case files would be quite labor intensive. There is generally no definitive statistical evidence that any of these factors affect price, but it is certainly plausible and no inconsistent with ethnographic reports. A great unknown is how much of the apparent stochastic variability in price and purity might be explained if data were available on all of these covariates. One of the advantages of getting multiple price observations self-reported by the same individual buyer is the ability to control for such covariates, e.g., through a fixed-effects regression.

A related issue is the possibility of differences between undercover buys made by a law enforcement agent and buys made by a “confidential informant” (CI) working for law enforcement. (The typical CI is a lower-level seller who has been arrested and is now cooperating with law enforcement in an effort to get a reduced sentence.) For instance, narcotics agents with local police have told the author that they know that CI’s
sometimes keep some of the PE/PI money. The agents might give the CI $50 to make a purchase from a certain drug house. The CI might pocket $20, buy $30 worth of drugs, and hand those drugs to the agents. Unless the CI is wired (risky and complicated), it is very difficult to stop this behavior and often what matters is evidence that some drugs were sold from the house, not whether it was 0.1 or 0.2 grams.

There might plausibly also be differences between buys made for purposes of making a case and those collected for intelligence purposes only, notably for the DEA’s Domestic Monitor Program. Indeed, Horowitz (2001) documents such differences, and Manski et al. (2001) suggest that intelligence-driven buys yield superior price information and recommend systematic collection of price data for research purposes. Even leaving aside human subjects consideration (is it ethical to collect data in a manner that intrinsically involves placing those collecting the data in physical jeopardy) the differences between intelligence-driven and enforcement-driven data might not necessarily imply the former are superior for research purposes. Law enforcement agents often display great courage and dedication in their efforts to get drug dealers off the streets, but given prevalent attitudes among law enforcement practitioners about the value of academic research, they might plausibly make buys that are for purposes only in a non-representative manner that reduces their personal risk of injury.

As a final note, STRIDE contains only samples that were analyzed in DEA forensic laboratories. Nationally, state and local police agencies make far more undercover purchases than do federal agents. A very useful next research project would be obtaining price data for a particular city that sends its samples to a good forensics laboratory and examining price variation and trends within that city at a finer unit of resolution than can be accomplished with STRIDE data.

5.0 Summary

The premise of this paper is that data on illicit drug prices and purity, imperfect though they are, can be valuable for a variety of research purposes, notably (1) understanding how prices affect use, (2) understanding how supply disruptions affect price, and (3) using price and purity data as “conversion factors” to help reconcile other drug-related aggregate measures in order to weave a holistic or systems understanding.
Unfortunately, drug price and purity data have a moderate number of idiosyncrasies that need to be appreciated in order for them to be used responsibly.

One set of issues pertains to the nature of the illicit drug markets themselves. For example, amounts paid need to be standardized for quantity discounts and quality premia, and this needs to be done in a manner consistent with the fact that drugs are an “experience good”.

The second set of issues pertains to the nature of the data and their collection. There are broadly three sources of price and purity data: (1) synthesis reports by government agencies, (2) self-report, and (3) transaction-level data from law enforcement. The last, particularly the DEA’s STRIDE database, is the most commonly used today, but it is by no means the only potential source of price and purity data, and the various sources have distinct sets of advantages and disadvantages.

Acknowledgements:
This paper is grounded in an understanding of drug markets and related data indicators that has evolved over the last 15+ years through many multi-lateral conversations with many drug policy analysts notably Mark Kleiman and a number of people affiliated in one way or another with RAND’s Drug Policy Research Center, notably Peter Reuter, Rosalie Pacula, and Rob MacCoun. At this point it is not so easy to identify and credit individual people with individual ideas, so I think the group collectively and take personal credit for all errors and omissions.

References:


