Estimating Cannabinoid Mass Sold from Noisy Cannabis Sales Data

ADAI | ALCOHOL & DRUG ABUSE INSTITUTE

Jason Williams, PhD[1], Dan Burgard, PhD[2], and Caleb Banta-Green, PhD, MPH, MSW[1]
[1]Alcohol & Drug Abuse Institute, University of Washington, [2]Chemistry Department, University of Puget Sound

Introduction

Passage of Washington State Initiative 502 (I-502) in November of 2012 initiated the difficult process of creating a regulated, functioning market where no (legal) market had existed before. I-502 thrust the former state Liquor Control Board, rebranded as the Liquor and Cannabis Board, into multiple new roles. Whereas the LCB had decades of experience in managing the state's liquor market in keeping with a robust set of federal regulations and requirements, the cannabis market had to be regulated *against* federal rules, with no institutions upon which to rely. Rules for product packaging, labeling, testing, distribution, and marketing had to be developed. Some direction (and legal cover) was provided by the federal Cole Memo in 2013, providing strong motivation to create a tightly regulated and enforced market to minimize diversion of products outside the legal market and/or to minors (Cole, 2013; Wikipedia, n.d.).

A major part of this regulation and enforcement is the state's seed-to-sale traceability database. This set of linked tables allows tracking of seedlings to final sale and from edibles and extracts back to original strains and sources. The system connects information from producers and processors, testing laboratories, and retailers "to prevent diversion and to promote public safety" (Washington State Liquor and Cannabis Board, n.d.) and, of course, facilitate tax collection. Quality assurance regulations require reporting by certified testing labs of the cannabinoid content (specifically, tetrahydrocannabinol [THC], tetrahydrocannabinolic acid [THCA], cannabidiol [CBD], and cannabidiolic acid [CBDA]) of each sampled batch of flower or refined product (Washington State Legislature, 2017). Thus, records combine product amounts with product potencies for every retail transaction. Knowing the amount of a given cannabinoid sold in a given area at a given time opens up a host of research questions. Our own investigation of these data is motivated by wanting to compare the THC sold to the THC metabolite found in wastewater testing over time, spanning implementation of I-502, to investigate the extent to which the legal market displaced the illegal market. Relating cannabinoids sold to key outcome variables would allow further investigation of associations with traffic accidents and fatalities (e.g., Huestis, 2015) and other adverse health effects (e.g., Hall, 2015). Comparing cannabinoids actually sold would further detail the extent to which cannabis is substituting for other substances such as alcohol (Anderson, Hansen, & Rees, 2013) and opioids (Bachhuber, Saloner, Cunningham, & Barry, 2014). Reliable data on price and potency for all product types would permit thorough investigation of the price elasticity of cannabinoids and the extent to which consumers are moving to more potent purchases (Smart, Caulkins, Kilmer, Davenport, & Midgette, 2017; Williams, Banta-Green, & Burgard, 2017).

As we have noted (Williams et al., 2017), the initial iteration of the Washington seed-to-sale database fell short of its promise for supplying usable sales data with potencies. In this paper, we briefly document some of those shortcomings and discuss a general framework for how one would create an estimate---no longer a simple accounting---of the amount of cannabinoids sold. We demonstrate a simple version of this framework to create a set of estimates for THC sold in King and Pierce Counties.

Calculating THC Sold: What We Need

Put simply, for every sale of a given product, we need to know the amount of that product sold and the density of THC per unit of product. Unlike alcohol, which is generally sold in liquid form labeled with an alcohol-by-volume potency that gives the percentage of each liter of liquid that is pure alcohol, cannabis products come in many forms: loose or combined flower, pre-rolled joints, oils and tinctures, sprays, various liquid and solid edibles, creams, bath salts, vape cartridges, and others to come in a growing and diversifying market. Amount of product sold might thus be expressed in terms of weight, volume, or units/servings per container. Density of THC would naturally have to match the form of the product: weight per unit weight, weight or volume per unit volume, or weight per unit.

Accounting for THC sold thus requires simply identifying relevant sales (e.g., based on geography and time), calculating the amount of THC in each transaction, and summing across sales:

$$\sum_{i=1}^{n} \text{Package Count}_i \times \text{Packages Sold}_i \times \text{Potency Per Count}_i \times \text{Scaling Factor}_i$$

where Package Count is the number of units per package, Packages Sold is the number of those packages in sale *i*, Potency Per Count is the density of THC per unit, and the Scaling Factor converts the result into the desired unit (e.g., mg or g). This more flexible form allows for 16 ounce or liter bottles of liquids, single- or 20-packs of edibles, etc. For products such as usable bud sold simply by weight, this simplifies to

$$\sum_{i=1}^{n} \text{Grams Sold}_i \times {}^{\text{mg THC}}/_{\text{g Product}_i}$$

where the units reflect the custom of expressing THC in milligrams and product sold in grams.

Calculating THC Sold: What We Have

The first version of the seed-to-sale database, implemented by BioTrackTHC (tables joined and flattened by Looking Glass Analytics to provide the data explored herein), divides products into the following categories: Solid Marijuana Infused Edible, Liquid Marijuana Infused Edible, Marijuana Extract for Inhalation, Marijuana Infused Topicals, Usable Marijuana, Marijuana Mix Package, and Marijuana Mix Infused. It provides fields for Packages Sold and Grams Sold. The former is used with the two edibles categories and topicals, the latter with concentrates, usable marijuana, and the mix categories. The database provides four potency fields: THC, THCA, Total THC, and CBD.

There is no package count. The following examples in Table 1 illustrate why this is a problem for edibles. Regulations limit edibles to 10mg THC per serving and 100mg per package (with some acceptable rounding). For the Oakor Slips, we have no idea whether this sale represents five single packs or five 20-packs, and thus whether this sale represents 22.9mg or 458mg or something in between. Assuming, again, that the potency provided is per serving, it appears from the product names that the two Dank Drizzle examples represent 10- and 20-serving bottles.

PackagesSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
5	4.58	4.36	4.43	0.24	Oakor 1:1 Slips	1:1 Slips
1	0.9	0.9	0.85	0	EH 4.20 Bar 3pk CBD Dark Choc. 30mg	RCC010D Dark Choc CBD 3-Pack
1	0.91	0.91	0	0	RCZ010 Dark Choc Hazelnut	RCZ010 Dark Choc Hazelnut
1	5.12	5.12	0.45	0	SM 50mg Dank Drizzle	Dank Drizzle
1	5.12	5.12	0.45	0	SM 100mg Dank Drizzle	Dank Drizzle

Table 1: Solid and liquid edibles illustrating lack of package count

There is no unit of measurement for the potency values. Note with the examples in Table 1 that we assumed potencies were mg per serving. Table 2 further illustrates how this makes calculating cannabinoid content difficult. The first two examples, branded ETHOS, are likely similar products but with wildly different reported THC potencies. For the Create spray, does the user get 0.95mg of THC per average spray? Does the entire 1 ounce bottle of Journey provide 78.7mg of THC, or is it 78.7% THC by weight or by volume? The relative similarity of reported CBD potency to the labeled potency of the Honey Lemon Dropperz and of reported THC for one of the two strawberry lemonades implies these potencies are indeed milligrams for the whole bottle.

PackagesSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
1	0.95	0.95	0	0	ETHOS Oral Spray Create	Create(10ml)
1	78.7	78.7	3.57	0	ETHOS Water Tincture Journey	Journey(1oz)
1	53.9	53.9	95.9	0	PTNZ - Dropperz - CBD Honey Lemon 100mg	PTNZ - Dropperz - CBD Honey Lemon 100mg
1	54.19	54.19	0	0	Strawberry Lemonade, 12 oz, 60 mg	Strawberry Lemonade, 12 oz, 60 mg
1	59.94	59.94	0	0	Strawberry Lemonade, 12oz, 80mg	Strawberry Lemonade, 12oz, 80mg
1	23.7	23.7	28.3	0	FAIRHAVEN RCBD1-60 12.19	FAIRHAVEN RCBD1-60 12.19
1	54.4	54.4	69.3	0	BALLARD D27-CBD 5.11	BALLARD D27-CBD 5.11

Table 2: Liquid edibles illustrating uncertainty about potency unit of measurement

Key database functions such as acceptable values and treatment of missing data appear to be lacking. In Table 3, the first entry probably accurately reports THC per item in the two-pack, but Total THC—defined in the Washington Administrative Code as 0.877×THCA + THC—has not been either entered or automatically calculated by the system. The second line, with all 0 entries, indicates the default value for each field in that particular lab's data entry system might be 0, which means any 0 value we find might not actually be 0. (Any entity can use any data entry system they wish, as long as they have the capability of transmitting data, as through an Application Programming Interface, to the LCB's system. There are actual missing potencies in the supplied database, implying at least some systems do capture and retain true missing.) The two Catapult Decaf sales illustrate a TotalTHC mis-entered or mis-calculated and another apparently representing THC + CBD. Either value seems unrealistically low for either a percent by weight or milligrams per serving. The Peanut Butter Cup also appears to have TotalTHC = THC + CBD, while the ZootBlast Single (which by regulation cannot be 30mg if actually a single

serving) appears to have TotalTHC = THC + THCA with some rounding error. Meanwhile, the reported THC for the Northwest Berry Lemonade is also so low as to not be appealing to the consumer but also not be quantifiable by a lab: Assuming potency was per serving with 10 servings to the bottle, that indicates 0.0112mg THC in 1.6oz liquid. Using the density of water (29.57g/oz), a lower bound on the density of the lemonade, this gives 0.0000112g ÷ 47.32g = 0.000024% by weight or 0.24 parts per million. Bonn-Miller and colleagues (Bonn-Miller et al., 2017) report a limit of quantification of 0.317% by weight, although labs in Washington claim 0.1% as the lower limit (Jikomes & Zoorob, 2018). Even if the potency was reported as grams per serving with 20 servings per bottle, the resulting density by weight appears to be lower than quantifiable by any lab.

PackagesSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
1	0	5.12	0	0	BLACKBERRY chocolate	10HBB - HYBRID Blackberry Choc 5mg x 2/pack (10mg)
1	0	0	0	0	FW Hot Cocoa Mocha	Catapult Hot Cocoa Java Single Serving
1	0.03	0.03	0	0	FW Hot Chocolate	Catapult Hot Cocoa Java Single Serving
1	0.006	0.06	0.05	0	Catapult Decaf Individual	Catapult Decaf Individual
1	0.11	0.06	0.05	0	Catapult Decaf Individual	Catapult Decaf Individual
2	0.92	0.91	0.01	0	Peanut Butter Cup	Peanut Butter Cup
4	0.98	0.95	0	0.02	ZootBlast - 30mg Single	ZootBlast - 30mg Single
1	0.0112	0.0112	0	0	OC SENSI Northwest Berry 100mg	Northwest Berry Lemonade, 16oz, 100mg

Table 3: More edibles illustrating unlikely potency values

The Marijuana Extract for Inhalation category, often referred to as concentrates, further illustrates these potency data problems in Table 4. Concentrates—vape cartridges, waxes, shatters, oils, resins, hashish, and kief—appear to be marketed by weight, even though many are liquids, with potencies as percent by weight (see, for example, Smart et al., 2017). If so, the two vape cartridges do not appear to be very concentrated. The kief appears to have had its THCA value entered in the CBD field. The Aliens on Moonshine has TotalTHC = THC + CBD, while the White Widow has TotalTHC = 0.877×THCA + THC + CBD. The last example appears to be a super-concentrated product, with more than 100% cannabinoid content. To be clear, the *assumption* is that potencies are consistently reported as percent by weight, despite these illustrations of inconsistency of reporting. A value that seems reasonable as a percent by weight for a concentrate—e.g., 75—could also be a reasonable value of milligrams THC in the package—e.g., a half gram vape cartridge (75mg THC ÷ 500g product = 15% THC). We have no way of knowing.

GramsSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
1	1.4	0.2	0	1.4	Royal THC Vape Pink Passion	Royal THC Vape Pink Passion
1	1.4	0.2	0	1.4	Royal THC Vape God Bud	Royal THC Vape God Bud
1	39.34	0.61	44.16	0	CCC KIEF 1	Kief
1.1556422	98.4	30.1	68.3	0	Aliens on Moonshine Clear	Aliens on Moonshine Clear
1	97.5	51.6	41.9	4.6	White Widow Clear	White Widow Clear
1	97.3	96.75	3.6	0.6	GG - Sugar Wax - 1.0 gram package	GG - Sugar Wax - 1.0 gram package

Table 4: Problematic concentrates examples

Topicals are troublesome but a small part of the market. The examples in Table 5 further illustrate potency problems described above. They also illustrate a problem heretofore apparent in the examples but not commented upon: The idiosyncratic nature of the two name fields, with apparently no guidance to producers and to retailers on what to put in these fields. This means the name fields are of limited utility in solving some of the mysteries documented here. Although package count/size and some indicator of potency are sometimes available within a sale, using product names to impute missing parameters for a given sale from other sales of the same product would be extremely difficult.

PackagesSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
2	0	0	0.45	0	Topical Flow Gel	Topical Flow Gel
1	0.008	0.08	0	0	Bath	Bath
1	0.99	0.53	0.01	0.36	Body Buzz(4oz)	Body Buzz(4oz)
1	0.99	0.53	0.01	0.36	Ethos Body Buzz Bath Salt 60mg	Body Buzz(4oz)
2	93.1	93.1	0.2	0	PTNZ - Lotionz - (5) Ampules 300mg	PTNZ - Lotionz - (5) Ampules 300mg
1	68.8	68.8	0.3	0	PTNZ - Lotionz - (5) Ampules 300mg	PTNZ - Lotionz - (5) Ampules 300mg

Table 5: Problematic topicals examples

The difficulty of relying on product name is further illustrated in Table 6, which includes problematic examples from the Marijuana Mix Package and Marijuana Mix Infused categories. These product types allow producers or retailers to combine cannabis from different lots or different plant parts, sell pre-rolled joints and other value-add products, and, in the latter category, add flavors and/or concentrates such as kief or hash oil. While some variation is to be expected in these categories, the two Third Kynde sales in Table 6 appear—if we trust the potencies—to be wildly different products in terms of cannabinoid profile. Meanwhile, the potencies for the Medi Haze product, nominally based on a strain marketed as having a CBD:THC ratio between 1:1 and 2:1, appear to possibly be entered on different scales, and it is unclear whence came the 56.5 Total THC for the Golden Nuggets.

GramsSold	TotalTHC	THC	CBD	THCA	RetailProductName	WholesaleProductName
0.88203125	9.27	1.7	6.73	8.63	Third Kynde - CBD Pre- roll	Third Kynde - CBD Pre- roll
0.82649842	12.7	12.7	18.7	0	Third Kynde - CBD Pre- roll	Third Kynde - CBD Pre- roll
3.5	2.54	0.11	0.16	2.26	Finished Flower - Medi Haze	Finished Flower - Medi Haze
1	56.5	10.5	4.2	37.5	Golden Nuggets No. 5	Golden Nuggets No. 5

Table 6: Problematic mixed package examples

In sum, the database provides no indicator of pack size or potency unit of measure. There appears to have been no specification to labs on how to enter potencies, such that even a potency that seems to be in the correct range could be representing a completely different value, such as percent THC by weight or by volume instead of milligrams of THC. There was no data checking to catch nonsensical values. Idiosyncratic descriptors mean we cannot always pull useful information out of names. While the system may work well for usable marijuana (and the very similar Marijuana Mix Package and Marijuana Mix Infused categories), there are clearly data entry errors and possibly 0 values that are actually missing values. Re-calculating Total THC as 0.877×THCA + THC will fix a number of errors, but this still relies on correct entry of the THCA and THC values.

A Way Forward

This exploration illustrates that the database is not adequate for a simple accounting of THC sold. Given the uncertainty, some outside information must be brought into the calculations to create an estimate instead. For example, regulations limit edible cannabis products to 10mg of THC per serving and 10 servings or 100mg of THC per package (Washington State Legislature, 2017). Many producers create servings of approximately 5mg per serving, encouraging users to start with one unit and to ingest a second if desired, and include 20 in a pack. This suggests pack size should be between 1 and 20.

All product types have suspicious potencies, from data entry errors if nothing else. For each product type, then, there needs to be some way to mark unlikely potencies. Low values, particularly those below 1% or 1mg, are unlikely except for products that are high in CBD. The data explored here were generated before the recent merger of the medicinal and "recreational" markets in Washington, and high CBD/low THC products are a very small part of the market represented in this data (Smart et al., 2017). Unfortunately, there is no indicator for whether a product is high in CBD or considered medicinal. As with pack size, product names inconsistently indicate potency or higher CBD content. We would expect some product categories to have higher THC potencies than others: Concentrates > Mix Infused > Usable Marijuana.

Our general framework, then, is to specify allowable high and low values, or brackets, for pack size and potency, by product type, to account for different assumed potency units of measure and different THC levels, and perhaps by high CBD products within each type. We assume the amount sold is accurately represented by either the Grams Sold field (for concentrates, usable marijuana, and the mix packages) or the Packages Sold field. (We set aside the idea that stores, receiving no instruction, sometimes use these fields differently, as when a concentrate sale has, for example, 2 entered for both Grams Sold and Packages Sold.) Our brackets designate the values we "trust", recognizing that without a great deal further investigation finding unlikely values within acceptable ranges, as when THCA potency is entered in the CBD field or vice versa, some of that trust is misplaced. Setting all values outside our brackets to missing, we then impute missing values from those within the brackets. Varying these assumptions allows creation of a set of estimates.

A Simple Approach

We demonstrate with a simplified version of this approach, in which we ignore high CBD sub-typing and use the median of "trusted" values to impute missing values. We make the following assumptions for each product type:

Solid edibles: This category appears to most often have some indicator of pack size—e.g., "2 pk", "5 serving"—in the product names. Pulling this information via regular expressions confirmed that pack counts range from 1 to 20, which we set as our bracket. For potency, there appears to be some acceptable violation of the 10mg per unit regulation due to variations in THC content of a given input strain and in the accuracy of the testing labs themselves (Jikomes & Zoorob, 2018). We use 12mg per unit as our top allowed potency, 1mg as our low end. Amount sold is calculated by multiplying our created pack size by the Packages Sold field.

Liquid edibles: While potencies around 5 are very likely per serving, and those above 50 likely for the whole bottle, judging which potency refers to per serving or per bottle in between is difficult: Should this cut-off be 10, even though we know potencies of 10% to 20% over the regulation are common? Should it be 12? Even if you could clearly mark a subset with per serving potencies, there is no number of servings per container by which to multiply it. The examples above illustrate that while a bottle size can sometimes be derived from the product names, servings per bottle is rarely if ever denoted. Potencies for liquid edibles are so problematic that we ignore the reported potencies entirely, derive milligrams of THC from the product names, and assume it applies

to the whole bottle, thus using Packages Sold as our amount sold. This a) ignores research about marketing not matching actual potency, and b) assumes the labeled potency refers to THC, when clearly this is not always the case. To counteract this slightly, we take the slightly conservative approach of using the 100mg maximum set forth in the regulations, and set any product marketed as "200mg" to 100mg of THC. We set the low end of the potency bracket at 9mg based on exploration of sales at the low end of estimated THC potency.

Concentrates: As noted, concentrates appear to be sold by weight, so we use the Grams Sold field (and ignore any weight information offered in the product names). Reported THC content appeared to have a maximum of 97.6% in the data, so we set our high value to 98% to account for any marginal extraction refinement. Investigation of products with lower THC values seemed to indicate legitimate values down to 2% THC in products with higher levels of CBD.

Topicals: These represent a tiny part of the market, with, again, no standardization of how potencies are reported. We ignore this product category.

Usable marijuana: Sold by the gram, reasonable potencies appear to go as high as 40% THC by weight. At the low end, investigation of products with reported potencies below 2% again found likely legitimate values given the CBD content, resulting in a lower Total THC bound of 0.6%.

Mix package: As noted, this category is quite similar to usable marijuana, so we adapt the same lowest allowed THC potency. On the high end, we allow for some creative blending and/or mis-categorization of product type and set the allowable potency at 50%.

Mix infused: Also sold by weight, we expect this product type to have higher THC levels than usable marijuana or the Marijuana Mix Package category. Exploration of the data indicated a bracket of 8% to 70% seemed reasonable.

Clearly each product type requires a certain set of assumptions, some stronger than others. Some of our choices were informed by exploration of the data, although even values that *seem* reasonable might be mistaken. Increasing the low end of a given bracket, say for usable marijuana, might exclude more values that are legitimately low because the cannabis strain is, for example, approximately 10:1 CBD:THC, and it might exclude more mistaken values and substitute the median instead. Again, without extensive research on expected potencies for a given strain and so on—noting that, like product names, there is no standardization of strain names—knowing which of these possibilities would predominate is impossible.

For any sales entry missing one or two of the required parameters, we plug in 1) the lowest possible value, 2) the middle possible value, the median of those values within the brackets, and 3) the highest possible value. We then calculate three values of milligrams of THC for each sales entry. For complete and "trusted" sales, these three values are identical. We then sum up across all sales in a given place and time to produce three estimated values of THC sold. We apply this approach to sales data from King and Pierce Counties for June of 2016.

Results

Table 7 summarizes the data quality with respect to our brackets for the King County data, and Table 8 repeats this for Pierce County. We specify the potency brackets in the tables, while the pack size bracket, 1–20, applies only to solid edibles. During June 2016, our one month test period, there were over 966,000 rows of data for King County, some of which were returns and some of which were daily summaries of sales for a specific product in a given store. Pierce County saw nearly 317,000 sales rows. We could derive pack counts for solid edibles and nominal potencies for liquid edibles from the product names for over half of each set of sales. Reported

potencies were by far more likely to be outside our brackets for solid edibles than for any other type (except, of course, liquids, for which we ignore reported potencies). Usable marijuana sales comprised the vast majority of rows of sales data, and only 0.3% of these sales had reported potencies either missing or below 0.6% or above 40% THC by weight.

Type	Rows	Bad pack count		Potency bracket	Bad THC	
Edibles	99092	41486	41.9%	1mg–12mg	23794	24.0%
Liquids	23153	0		9mg–100mg	10025	43.3%
Concentrates	119464	0		2%–98%	784	0.7%
Bud	684264	0		0.6%–40%	2328	0.3%
Mix package	23547	0		0.6%–50%	51	0.2%
Mix infused	16858	0		8%–70%	176	1.0%
Total	966378	41486	4.3%		37158	3.8%

Table 7: King County sales data quality

Type	Rows	Bad pack count		Potency bracket	Bad THC	
Edibles	19240	8768	45.6%	1mg–12mg	4389	22.8%
Liquids	5200	0		9mg–100mg	2378	45.7%
Concentrates	35148	0		2%–98%	281	0.8%
Bud	245934	0		0.6%–40%	665	0.3%
Mix package	8008	0		0.6%–50%	0	0.0%
Mix infused	3149	0		8%–70%	0	0.0%
Total	316679	8768	2.8%		7713	2.4%

Table 8: Pierce County sales data quality

Tables 9 and 10 report our estimate of low, middle, and high amount of THC sold by category and in total for each county in June 2016. Our estimate is that approximately 383,420 grams of THC were sold in King County, and 153,165 grams in Pierce County. As seen in the tables, in percentage terms our lowest possible value and highest possible value are not far apart, 3.3% of the middle value in King County and about 2% in Pierce County. In terms of milligrams, the high and low estimates are over 12kg and 3kg of THC apart, respectively. The high and low possible values are not far apart in percentage terms because usable marijuana, a product category for which we have little choice but to trust over 99% of sales entries, still dominated the market, accounting for approximately 80% of the THC sold in these two counties.

Type	Pot. bracket	Low THC (mg)	Mid THC (mg)	High THC (mg)	Mid % of total
Edibles	1mg–12mg	2,682,307	5,298,512	11,850,677	1.4%
Liquids	9mg–100mg	985,337	1,333,405	2,007,085	0.3%
Conc.	2%–98%	58,890,255	59,202,127	59,323,155	15.4%
Bud	0.6%–40%	303,650,863	304,619,042	305,551,430	79.4%
Mix package	0.6%–50%	7,761,970	7,786,011	7,824,807	2.0%
Mix infused	8%–70%	5,161,408	5,180,861	5,232,637	1.4%
Total		379,132,141	383,419,958	391,789,791	
	mg difference vs. middle	-4,287,817		8,369,833	
	% difference vs. middle	-1.12%		2.18%	

Table 9: King County THC sold estimate

Type	Pot. bracket	Low THC (mg)	Mid THC (mg)	High THC (mg)	Mid % of total
Edibles	1mg–12mg	492,576	780,643	2,410,903	0.5%
Liquids	9mg–100mg	202,213	286,750	450,370	0.2%
Conc.	2%–98%	21,391,868	21,528,010	21,578,828	14.1%
Bud	0.6%–40%	126,625,016	126,976,180	127,303,244	82.9%
Mix package	0.6%–50%	2,578,514	2,578,514	2,578,514	1.7%
Mix infused	8%–70%	1,014,812	1,014,812	1,014,812	0.7%
Total		152,304,999	153,164,908	155,336,671	
	mg difference vs. middle	-859,909		2,171,763	
	% difference vs. middle	-0.56		1.42%	

Table 10: Pierce County THC sold estimate

To provide a more direct comparison of consumption in the two counties, we provide crude milligrams per capita results in Figure 1. To reiterate, the upper and lower bounds presented here are not confidence intervals in the usual sense. They represent our calculation, given our assumptions, of the lowest possible and highest possible amount of THC sold. That said, the two counties appear to have purchased roughly the same amount of THC per person.

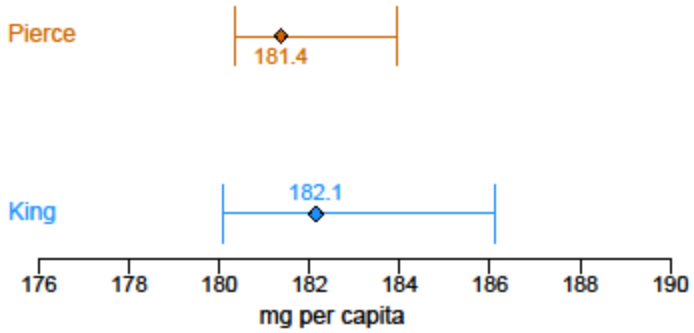

Figure 1: Comparing THC sold estimates accounting for population size

Discussion

While the first generation of the Washington seed-to-sale database, as implemented by BioTrackTHC, was functional for some users for some purposes, our focus here is on the seemingly simple issue of calculating cannabinoids sold across all sales in a given place and time. For that purpose, the data are problematic. There is no indicator of package size for edibles, only the number of packages sold. Guidance and/or over- sight on how to enter potencies may have been incomplete or ignored, such that some might be entered as proportions by weight or by volume, some might be entered as grams or other weight units of THC or CBD, some are for the whole package and some for a single serving, and "Total THC" sometimes includes CBD and other cannabinoids. Unrealistically low and unrealistically high potencies appear, illustrating a lack of data checking. The data systems of some labs may have default values of 0 for the potencies, such that missing values were transmitted to BioTrackTHC as 0. The extent of data entry errors, including both "fat finger" errors and putting the right value in the wrong field, is unknown but clearly well above 0. There is no standardization of names, which would aid in filling in missing information from similar sales and aid extraction of package size. Please note that our observations of flaws are related to assessing cannabinoid content only, and we cannot judge the extent to which the system was successful in "prevent[ing] diversion and. . . promot[ing] public safety" (Washington State Liquor and Cannabis Board, n.d.).

In order to still make use of these data, we proposed moving from calculation to estimation by making a set of assumptions about the underlying data. Changing these assumptions will result in marginally different results, as most of the market remains usable marijuana and the similar mix packages categories—i.e. products sold by weight where potency is recorded in the more traditional percent by weight. While these categories likely have random data entry errors, some of which may put the potencies outside of our brackets, their reported potencies are probably more accurate than the other categories for which confusion about potency units of measurement are more likely (to say nothing of pack size).

Several refinements could improve estimation. First, defining high CBD products within each product type would allow for higher and lower THC potency brackets. There is no high CBD or "medicinal" indicator in these data, and defining it based on reported CBD, of course, requires accurately reported CBD. Second, more elaborate data checking algorithms could flag items with logically inconsistent reported potencies, as with our Sugar Wax (Table 4) with more than 100% cannabinoid content by weight (assuming, again, those potencies were supposed to be percent by weight). Third, Monte Carlo simulation, drawing on likely distributions of values either missing or outside of acceptable bounds, could produce 95% confidence intervals. Given that our simple approach produced upper and lower bounds less than 3.5% apart, such refinements would, again, have marginal benefit.

References

Anderson, D. M., Hansen, B., & Rees, D. I. (2013). Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption. *Journal of Law & Economics*, *56*(2), 333–369.

Bachhuber, M. A., Saloner, B., Cunningham, C. O., & Barry, C. L. (2014). Medical Cannabis Laws and Opioid Analgesic Overdose Mortality in the United States, 1999-2010. *JAMA Internal Medicine*, *174*(10), 1668–1673. http://doi.org/10.1001/jamainternmed.2014.4005

Bonn-Miller, M. O., Loflin, M. J. E., Thomas, B. F., Marcu, J. P., Hyke, T., & Vandrey, R. (2017). Labeling Accuracy of Cannabidiol Extracts Sold Online. *Journal of the American Medical Association*, *318*(17), 1708–1709. http://doi.org/10.1001/jama.2017.11909

Cole, J. M. (2013). Guidance for Marijuana Enforcement. Washington, DC: US Department of Justice. Retrieved from https://www.justice.gov/iso/opa/resources/305201382913275685746 7.pdf

Hall, W. (2015). What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? *Addiction*, *110*(1), 19–35. http://doi.org/10.1111/add.12703

Huestis, M. A. (2015). Cannabis-Impaired Driving: A Public Health and Safety Concern. *Clinical Chemistry*, *61*(10) 10–12. http://doi.org/10.1373/clinchem.2015.245001

Jikomes, N., & Zoorob, M. (2018). The Cannabinoid Content of Legal Cannabis in Washington State Varies Systematically Across Testing Facilities and Popular Consumer Products. *Scientific Reports*, *8*(1), 4519. http://doi.org/10.1038/s41598-018-22755-2

Smart, R., Caulkins, J. P., Kilmer, B., Davenport, S., & Midgette, G. (2017). Variation in cannabis potency and prices in a newly legal market: Evidence from 30 million cannabis sales in Washington state. *Addiction, 112*, 2164--2177. http://doi.org/10.1111/add.13886

Washington State Legislature. Marijuana Licenses, Application Process, Requirements, and Reporting, Pub. L. No. 314–55 (2017). Washington Administrative Code 314-55. Retrieved from http://app.leg.wa.gov/WAC/default.aspx?cite=314-55

Washington State Liquor and Cannabis Board. (n.d.). Traceability Reporting. Retrieved March 14, 2018, from https://lcb.wa.gov/mjtrace/reporting

Wikipedia. (n.d.). Cole Memorandum. Retrieved from https://en.wikipedia.org/wiki/Cole_Memorandum

Williams, J., Banta-Green, C., & Burgard, D. (2017). Commentary on Smart et al (2017): The need for better marijuana sales data. *Addiction*, *112*(12), 2179–2180. http://doi.org/10.1111/add.14037

Citation: Williams JR, Bugard D, Banta-Green CJ. Estimating Cannabinoid Mass Sold from Noisy Cannabis Sales Data. Alcohol & Drug Abuse Institute, University of Washington, June 2018.
URL: http://adai.uw.edu/pubs/pdf/2018estimatingcannabinoidmasssold.pdf.

See related PowerPoint presentation "Estimating cannabinoids sold from marijuana sales data in Washington State: Watch me pull a number out of my hat" by Williams JR et al. March 2018.
URL: http://adai.uw.edu/pubs/pdf/2018THCdifficultiesandinitialsimpleestimate.pdf

This report was produced with support from the Washington State Dedicated Marijuana Fund for research at the University of Washington.