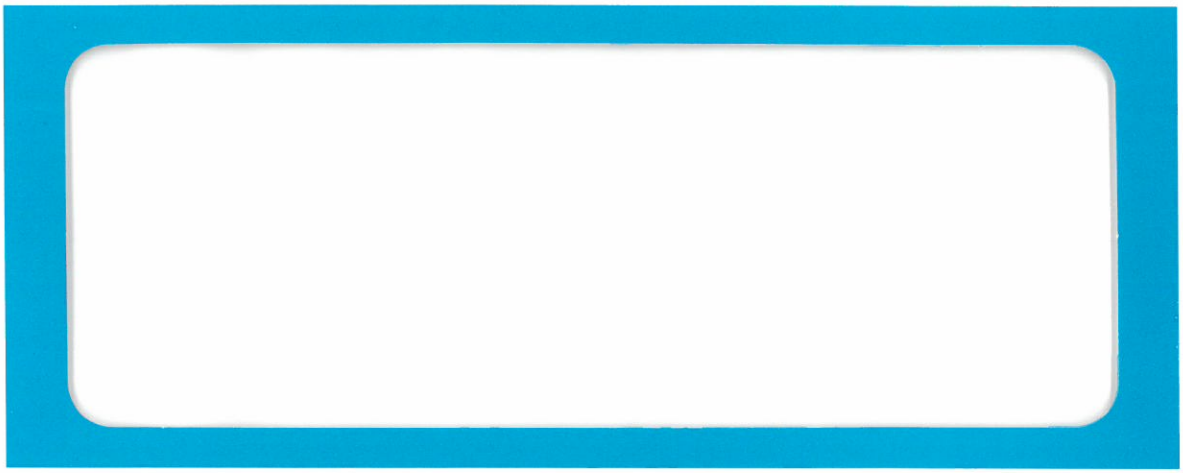
Methodology Branch

Direction de la méthodologie



Household Survey Methods Division

Division des méthodes d'enquêtes auprès des ménages

11-619E/F
no. 2013-005
c. 2

Statistics Canada Statistique Canada

Canada

WORKING PAPER
METHODOLOGY BRANCH

Updating the Police-Reported Crime Severity Index Weights: Refinements to the Methodology

HSMD-2013-005E/F

Colin Babyak, Aidan Campbell, Rose Evra and Sarah Franklin

Household Survey Methods Division
Statistics Canada
July 2013

Updating the Police-Reported Crime Severity Index Weights: Refinements to the Methodology

Abstract

Statistics Canada produces three principal measures of crime, the police-reported crime rate (from the Uniform Crime Reporting Survey (UCR)), the victim-reported crime rate (from the General Social Survey (GSS)) and the crime severity index (CSI) from administrative justice data sources. In order to calculate the CSI, each violation is given a crime severity weight. The crime severity weights are calculated by multiplying the average length of incarceration and rate of incarceration for a given offence. Every five years, the crime severity weights are reviewed to reflect changes in sentencing patterns and legislation. Since the weights were first calculated in 2008, 2013 will be the first time they have been updated. While updating the weights, the strategy used to calculate them was slightly refined. This document will describe the refinements in the methods used to produce the new crime severity index weights.

Key Words: Volume Index, Crime Measurement, Crime Severity Index Weights

Résumé

Statistique Canada produit trois mesures clés de la criminalité – le taux de crimes déclarés par la police (d'après le Programme de déclaration uniforme de la criminalité (Programme DUC), le taux de crimes déclarés par les victimes (à partir de l'Enquête sociale générale (ESG)), et l'Indice de gravité de la criminalité (IGC), établi à partir de sources de données administratives sur la justice. Aux fins de calculer l'IGC, chaque infraction se voit attribuer un poids correspondant à sa gravité. Les poids sont obtenus en multipliant la durée moyenne d'incarcération et le taux d'incarcération pour une infraction donnée. Tous les cinq ans, ces poids sont examinés de manière à rendre compte des changements touchant les pratiques de détermination de la peine et des modifications apportées aux lois. Étant donné que ces poids ont été calculés pour la première fois en 2008, 2013 marque leur première mise à jour. Dans le contexte de cette mise à jour, la stratégie utilisée pour calculer les poids a été légèrement améliorée. Le présent document décrit les améliorations apportées à la méthode servant à produire les nouveaux poids de l'Indice de gravité de la criminalité.

Mots clés : indice de volume, mesure de la criminalité, poids de l'Indice de gravité de la criminalité

1. Introduction

Statistics Canada produces three principal measures of crime, the police-reported crime rate (from the Uniform Crime Reporting Survey (UCR)), the victim-reported crime rate (from the General Social Survey (GSS)) and the police-reported crime severity index (CSI). In order to calculate the CSI, each violation is assigned a weight which is calculated by multiplying the average length of incarceration and rate of incarceration for a given violation. The crime severity index weights are a tool to calculate the weighted change in the volume of crime. Every five years, the weights are reviewed to reflect changes in sentencing patterns and legislation. Since the weights were first calculated in 2008, 2013 will be the first year they have been updated. While updating the weights, the strategy used to produce them was reviewed. This document describes refinements in the methodology used to determine the new crime severity index weights.

As of January 1st, 2013, there were 220 UCR violations requiring a weight. When the weights were first created, in 2008, there were 150 violations. Even after a violation has expired (i.e., due to legislative changes), a weight needs to be calculated since some police incidents are reported to the UCR survey using the old violation.

The Integrated Criminal Courts Survey (ICCS) data are used to calculate the weights.

1.1 Overview of the Crime Severity Index

The CSI is a volume index, in that it is principally intended to measure changes in volume over time, as opposed to the Consumer Price Index (CPI), which is intended to measure price changes over time (Statistics Canada, 2008). Whereas the traditional crime rate is simply the sum of incidents of criminal violations taking place in a year divided by the population (Statistics Canada, 2003), the CSI adds to this notion by weighting each incident by a measure of relative severity for the violation and then dividing by the population (Babyak et al., 2009). For the purposes of calculating the overall crime rate, for example, one incident of murder is equivalent to one incident of disturbing the peace. In the CSI, however, an incident of murder has a weight over 7,000 while an incident of disturbing the peace has a weight of just ten.

The CSI weights are calculated at the national level. This permits comparisons between provinces and any geographical region in the country. Every crime has the same weight throughout the country; a level 1 assault in Ontario is considered the same as a level 1 assault in Nunavut.

The CSI is calculated by multiplying the number of incidents of a violation by the weight for that violation, summing them up and then dividing by the population. This is standardized by dividing by the baseline data, namely the weighted sum for Canada for the base year, and multiplying by 100. The equation for the national CSI is given in (1.1).

$$CSI_{t/b} = \frac{\sum_{i=all\_offences} q_{i,t} w_{1i} / pop_t}{\sum_{i=all\_offences} q_{i,b} w_{1i} / pop_b} * 100 \qquad (1.1)$$

The way this equation reads is: the police-reported Crime Severity Index for time period t, given base year b, is equal to the weighted sum of all incidents of a violation in time period t divided by the weighted sum of all incidents of a violation in time period b, multiplied by 100. The population figures are included to standardize the CSI, since crime volume is directly dependent on the size of the population.

The CSI weights need to be updated periodically. One reason for this is that new crime legislation may generate a new violation for which a weight must be calculated. Also, changes in court sentencing practices over time may change the ranking of the severity of crimes. In the case of the CSI, updating the weights every five years seemed most appropriate, especially since the weights use five years' worth of courts data.

The updating strategy incorporates a backwards-linking chain index. Essentially, in the linkage year, denoted LY in (1.2) and (1.3), the index is calculated with the old weights and the new weights, using the same crime volumes for the linkage year. This ratio, calculated nationally, is defined as the linkage factor and is applied to every CSI value that is calculated using new CSI weights. Using a linkage factor ensures comparability of CSI values using new and old weights.

$$CSI_{t/b,p} = \frac{\sum_{i=all\_offences} q_{i,t,p} w_{2i} / pop_{t,p}}{\sum_{i=all\_offences} q_{i,b} w_{1i} / pop_{b}} * LF_{LY} * 100 \tag{1.2}$$

$$LF_{LY} = \frac{\sum_{i=all\_offences} q_{i,LY} w_{1i}}{\sum_{i=all\_offences} q_{i,LY} w_{2i}} \tag{1.3}$$

In order to calculate a regional CSI, the numerators in formulae 1.1 and 1.2 are calculated with respect to the region and the denominators with respect to all of Canada. Thus, a region's CSI for a given year is relative to the national 2006 base year. The linkage factor is always calculated nationally. The formulae for calculating the index across different geographies and multiple linkage factors can be found in the official specifications for the index (Babyak, 2008).

2. Refinements to the CSI Weighting Methodology

The methods used to calculate the CSI remain largely the same as before, but parts of the methodology used to calculate the weights were updated to improve the quality of the weights. The following describes these refinements.

2.1 Outlier Detection

During the calculation of the CSI weights, some court sentences for specific charges appeared to be outliers. An investigation was undertaken to determine what the causes of these outliers were and what treatment they should receive. It is worth noting that the courts program at Statistics Canada processes more than ten million court appearance records per year. When rolled up to the charge level, the number of records used to calculate the weights, over five years, for the CSI weights, included some three million guilty charges. While examining outliers by violation code[1], some unique cases were discovered. The sheer volume of court records makes it impossible to follow up every questionable sentence and thus it is believed that the majority of the outliers in the courts data are due to data quality issues as opposed to unique but valid charges. Regardless of the reason, courts sentencing data has proven to be of very good

[1] A violation code is a 4-digit code that categorizes criminal offences into 220 categories.

quality with few issues such as that above. For the revised CSI weights, only 206 charge records between 2006 and 2010 were identified as outliers.

Prior to calculating the average length of incarceration, outlier detection is performed, by UCR violation. Since there is a significant difference in the sentencing patterns for adults and youths, the outlier detection is conducted separately for adult and youth data. Once the outliers have been detected and removed, the remaining adult and youth data are combined to calculate the average length of incarceration for each violation. It should be noted that outliers are only eliminated from the average length of incarceration, not the incarceration rate (these offenders were sent to custody, it is only their length of custody that is outlying).

When the CSI weights were first calculated five years ago, the outlier detection method used quartiles of log-transformed lengths of custody, with mainly short lengths of custody being identified as outliers. For the revised CSI weights, following some investigations, it was determined that short lengths of custody should not be flagged as outliers; they are potentially all valid. One day sentences are common since the variable used to determine the length of custody (LCUST) represents only the time to be served after sentencing. Time served prior to sentencing is not captured causing the data set to include one day sentences.

For the revised CSI weights, in order to avoid eliminating short lengths of custody while still identifying extremely high lengths of custody, the non-parametric sigma-gap method (Bernier and Nobrega, 1998) is used. This method was developed by Statistics Canada in 1983. It consists of ordering the data set, by length of custody for each violation, and then calculating the distances between adjacent pairs. In the case of the CSI, if the difference between two lengths of custody is greater than three standard deviations, then lengths of custody beyond the gap are flagged as outliers (i.e., if $x_{i+1}-x_i > 3\sigma_x$, then all observations greater than x_i are outliers). The results of the sigma-gap method were validated by inspecting graphs of lengths of custody for each violation: the sigma-gap method identifies records which visually appear to be outliers. Often, it identifies life sentences for non-homicide violations. It also identifies very few outliers.

2.2 Unknown Sentences

To obtain the rate of incarceration per violation, the ICCS microdata (both adult and youth) for the time period under consideration are used. Only records with known sentences are kept. As a result of changes in the way data are extracted from the courts' record management systems, Statistics Canada receives more records with unknown sentences than it did when the CSI weights were first calculated. Consequently, the identification of charge records with unknown sentences was scrutinized and refined for the new CSI weights.

2.3 Minimum Charges

A minimum number of charges is required to calculate the weight for a violation. For violations with less than five guilty court charges a proxy weight is assigned. In addition to the minimum number of charges, at least one charge must have a length of custody greater than one day. A proxy weight is produced in these cases since a weight of zero for a crime would be nonsensical (the crime would never contribute to the CSI) and a weight of one or less would not adequately reflect the nature of sentences for that violation.

3. Record linkage

In some instances, the offence information on the 2006-2010 ICCS data was not detailed enough to accurately identify the UCR violation. In order to supplement the information available in the courts' data used to estimate the weights, a record linkage of court data with UCR data was performed for some violations.

3.1 Theft

Up until 2011, the courts data did not indicate if the theft charges were for theft of a motor vehicle, they only indicated if the theft was over $5,000 or under $5,000 (UCR violation codes 2130 and 2140), even though two UCR codes were dedicated to motor vehicle theft (2131- theft of a motor vehicle over $5,000, and 2141 - theft of a motor vehicle under $5,000). In April 2011, a new UCR code for motor vehicle theft was introduced (2135) to coincide with new legislation, at which time codes 2131 and 2141 were expired.

In order to identify court charge records involving theft of a motor vehicle, record linkage was performed. Data from police incident files were merged with the ICCS data to identify motor vehicle theft according to the police.

More than 74% of the theft charge records linked to a UCR incident. Deterministic record linkage was performed. The variables used to link courts and policing data were: the offender's soundex[2], date of birth, sex, the jurisdiction where the incident occurred, and the date the criminal incident occurred.

For charge records which did not link to a police incident record, motor vehicle theft was imputed using the ratio of motor vehicle theft that was determined by the linked charge records. The imputation distinguished between charge records where the person was sentenced to custody and those who were not. Using these ratios, a motor theft (2135) was randomly assigned to each remaining unlinked theft charge.

3.2 Drugs

A similar record linkage was performed for drug violations. In the ICCS data, drug related charges falling under the Controlled Drugs and Substances Act are all labelled as 'other' type of drug. Consequently, the charge records only distinguish between possession, trafficking, importation and exportation, and production of drugs. The UCR violations, however, distinguish between various types of drugs: heroin, cocaine, cannabis, methamphetamine (Crystal Meth), methylenedioxyamphetamine (Ecstasy), and 'other drugs'. The sentencing pattern is quite different depending on the type of drug. Also, the proportion of drug violations by type of drug is quite different in the police data than in the courts data. The courts have a higher proportion of more serious drugs whereas police have a higher proportion of less serious drugs, mainly cannabis. Consequently, it was important to be able to calculate CSI weights by drug type.

Similar to motor vehicle theft, record linkage was performed between policing and courts' data in order to assign the drug type according to the police.

More than 80% of the court drug charges linked to a UCR incident. Deterministic record linkage method was performed using the same variables that were used for motor vehicle theft: the offender's soundex, birthdate, sex, the jurisdiction where the incident occurred, and the date the incident occurred.

[2] The Russell Soundex System is a phonetic algorithm that encodes names that sound the same in English.

For the 20% of drug charges that did not link to a police incident, a drug type was imputed. Imputation was performed by type of violation (i.e., possession, trafficking, etc.); whether the person was sentenced to custody or not; and the length of custody. Once these imputation classes were created, ratios for each type of drug were calculated. The linked charges were used to determine the ratios. Using these ratios, a type of drug was randomly assigned to each remaining drug charge (more details are given in Evra, Franklin, Babyak and Campbell (2013)).

4. Proxy Weights

There is not sufficient data to calculate a CSI weight for every violation. Some crimes are very rare and other violations are new (the Criminal Code is regularly updated with new legislation). As a result, even though the police can arrest someone for a crime, the ICCS may not have any court sentencing data for that violation. The need thus exists to calculate proxy CSI weights. As one of the objectives of the CSI is to remove as much as possible the element of subjectivity, this was not a straightforward task. The following is a description of the different methods used to calculate proxy weights.

4.1 Criminal Code sections map to multiple UCR violation codes (Group 1)

A small number of Criminal Code sections map to more than one UCR violation code. For the weight calculations, these are treated as one violation and the same weight is assigned to all violations derived from the same portion of the Criminal Code. For example, while Break and Enter (2120) has its own section under the Criminal Code, Break and Enter to steal a firearm (2121) and Break and Enter a Motor Vehicle to steal a Firearm (2125) fall under the same section of the Criminal Code and thus receive the same CSI weight.

4.2 Insufficient court charge records (Group 2)

The following is a description of the proxy weight calculation process when there are not sufficient court charges for a violation to calculate an accurate weight. Any UCR violation with the same first three digits as the violation code in question and with the same UCR seriousness ranking[3], are grouped together and a proxy weight is calculated from these violations. If no such violation exists then violation codes with the same first two digits and the same seriousness ranking are grouped together, and the process repeated. Again, if no such code exists then those with the same first digit are grouped, and the process repeated.

If there are no violations with the same seriousness ranking then violation codes with the closest seriousness ranking are examined. Of all the violations with the closest seriousness rank, those which share the first three digits with the code in question are examined. If there are none, then those with the same first two digits are examined. If none exist, then violation codes with the same first digit are used.

4.3 Insufficient length of custody (Group 3)

In addition to insufficient court charges, if the only sentences for a violation are zero or one day sentences, then the average sentence time would be between 0 and 1 day. It was felt that this would not accurately reflect the average custody sentence. Consequently, proxy weights will also be calculated for any violation in this situation; they will be assigned the minimum weight of similar violations (i.e., same first UCR2 digit and same or closest rank).

[3] Seriousness ranking is based on the maximum penalty for a given violation and the type of violation. So, two violations against the person with a maximum penalty of 25 years would have the same seriousness ranking.

4.4 Special cases

In two special cases, proxy weights were required for which the previous proxy rules could not be applied. Subject matter experts were consulted to determine the most appropriate methodology. This is described in detail below.

Other Sexual Crimes (1340)

The UCR code for Other Sexual Crimes (1340) expired in 2008 due to the introduction of more detailed UCR codes. However, since some police incidents are still being coded to 1340, a weight for this expired code was needed. In order to produce a weight similar to the one that was calculated when the CSI was first designed, the weight for violation 1340[4] was calculated based on data from all charges that would have been formerly coded to 1340.

Terrorism (3711 to 3717)

There were few court charges with violations relating to terrorism, making it impossible to calculate a weight for each of these violations. The generic approach to assigning a proxy weight lead to weights that were deemed too low. Also, a concern was raised because the terrorism violation codes begin with "37" like many codes of unrelated violations which would have been used to determine the proxy weight. Consequently, to calculate the new CSI weights, all of the terrorism violations were combined to produce a single weight for the group. That weight was then assigned to all terrorism violations.

4.5 From proxy to calculated weight

Many UCR violations have been created since the CSI weights were first calculated. These new violations were assigned proxy weights the year they were introduced to the UCR. With the new CSI weights, using courts data from 2006 to 2010, there is now enough court data to calculate non-proxy weights. A total of 28 UCR codes went from proxy to non-proxy weights with the new CSI weights (listed in table 1 below).

While many high volume offences were seen to have fairly accurate proxy weights, others saw their weights change dramatically in the shift from proxy to calculated weights. This illustrates the need for the CSI weights to be updated through the years to capture changes to the Criminal Code and UCR violations.

[4] The following violations would have been coded to 1340 prior to 2008: 1345, 1350, 1355, 1356, 1360, 1365, 1375 and1380.

Table 1: UCR violation codes which no longer have proxy weights

Name	Code	# charges	Original CSI weight (proxy)	Revised CSI weight (not proxy)
Sexual Interference (effective 2008-04-01)	1345	5,127	211.0	312.3
Invitation To Sexual Touching (effective 2008-04-01)	1350	1,831	211.0	380.8
Sexual Exploitation (effective 2008-04-01)	1355	953	211.0	486.3
Sexual Exploitation of a Person with a Disability (effective 2008-05-01)	1356	6	294.6	376.5
Incest (effective 2008-04-01)	1360	248	678.3	881.4
Corrupting Children (effective 2008-04-01)	1365	76	294.6	314.5
Luring a Child via a Computer (effective 2008-04-01)	1370	431	171.9	368.7
Anal Intercourse (effective 2008-04-01)	1375	66	211.0	718.6
Bestiality / Commit / Compel / Incite a Person (effective 2008-04-01)	1380	31	211.0	118.0
Voyeurism (effective 2008-04-01)	1385	383	85.5	42.1
Using Firearm/Imitation of Firearm in commission of offence (eff. 2008-04-01)	1455	2,253	267.4	322.2
Pointing a Firearm (effective 2008-04-01)	1457	1,285	194.0	209.8
Assault Against Peace Officer with a weapon or causing bodily harm(eff. 2009-10)	1461	49	79.4	116.4
Trap Likely To or Causing Bodily Harm (effective 2008-04-01)	1475	19	398.6	272.4
Kidnapping (effective 2010-01-08)	1515	38	477.4	1,410.3
Forcible Confinement (effective 2010-01-08)	1516	799	70.4	356.2
Trafficking in Persons (effective 2005-11-01)	1525	11	1,278.0	423.4
Intimidation of a Justice System Participant or a Journalist effective (2008-04-01)	1621	1,191	66.5	102.3
Indecent/Harassing Telephone Calls (effective 2008-04-01)	1626	4,245	17.3	19.5
Break and Enter to steal firearm (effective 2008-05-01)	2121	92	187.0	301.3
Identity Theft (effective 2010-01-08)	2165	265	48.2	144.5
Identity Fraud (effective 2010-01-08)	2166	3,132	87.4	88.3
Utter Threats Against Property or Animals (effective 2008-04-01)	3540	5,372	29.3	39.2
Public Incitement Of Hatred (effective 2008-04-01)	3560	14	29.3	54.7
Other Criminal Code Violations: Fail to stop causing bodily harm (effective 2011)	9312	45	153.6	169.1
Other Criminal Code Violations: Fail to stop (effective 2011)	9313	991	61.6	52.3
Dangerous Operation Causing Bodily Harm While Street Racing	9440	7	316.0	55.7
Dangerous Operation of Motor Vehicle While Street Racing	9450	19	24.0	17.3

5. Data Quality

The crime severity weights were calculated using ICCS data from 2006 to 2010[5]. More details on data quality evaluations are given in Evra, Franklin, Babyak and Campbell (2013). There were approximately 3 million guilty criminal charges in the dataset. Of that number, almost 1.3 million were charges where the offender was sentenced to custody and approximately 1.2 million were sentenced to a length of custody greater than zero days[6]. Only charge records with lengths of custody greater than zero were

[5] The ICCS uses fiscal year data, April 1st until March 31st of the following year. So, the 2006 data covers April 1st, 2006 to March 31st, 2007, and the 2010 data covers April 1st, 2010 to March 31st, 2011.

[6] No length of incarceration was available for Manitoba, all length of custody equal 0.

included in the average length of incarceration calculation, after the outlier detection step where 206 charges were flagged and removed as outliers.

It should be noted that the unit of analysis for the CSI weight calculations (i.e., incarceration rate, average sentence length) is the charge record. The ICCS data are available at the charge level. A given court case may involve multiple charges. The 3 million guilty courts charges between 2006 and 2010 represent 1.4 million court cases. Approximately 48% of these courts cases had more than one charge. Of the cases with multiple charges, 77% had the same sentence length for all charges where the person was sentenced to custody. For cases where all charges in the case have the same length of custody, it could be that this is actually case-level information repeated on each charge record, or these could be charges for which identical sentences were handed out, to be served successively or concurrently.

The ICCS data extraction process has changed since the first CSI weights were created. Because of concerns about how this could affect the CSI weights, a number of data quality analyses were performed to evaluate the charge-level data. These analyses are summarized below.

5.1 ICCS Decision Variable

The ICCS "decision" variable categorizes the different type of court verdicts. A value of 10 (found guilty), 11 (found guilty, but not convicted) and 20 (convicted of a lesser offence) is assigned to this variable when the suspect is found guilty of the charge.

After noticing a drop in the number of charge records with decision 11 in the 2006-2010 versus 2002-2006 ICCS data, a thorough review of this variable was undertaken. Only 0.55% of the charges were decision 11 or 20 in the 2006 to 2010 court dataset, whereas 14.5% of guilty charge records were decision 11 or 20 in the 2002 to 2006 court dataset used to calculate the original CSI weights. There was a change in the data reporting process between 2002 and 2010, specifically regarding how court data are extracted from a jurisdiction and sent to Canadian Center for Justice Statistics (CCJS). However, this change does not appear to have impacted the incarceration rate or average incarceration time.

5.2 Length of custody variable

The length of custody variable (LCUST) represents the number of days the guilty offender is sentenced to stay in custody following sentencing. Custodial sentence length excludes time spent in custody prior to sentencing and/or the amount of credit awarded for time spent in pre-sentence custody. Charge records for which the length of custody was unknown were excluded from the average length of custody calculation (section 2.2). The length of custody variable can take a value up to 9,125 days which is equivalent to 25 years (i.e., a life sentence).

Some offenders are sentenced to a single day, but the reasons behind such a sentence are varied. It could mean that the judge considers the time served prior to sentencing to be sufficient punishment for the crime committed, or that the offender needs to return to prison for a day for administrative reasons, or it could be a legitimate one-day sentence. There were a total of 180,938 custody sentences with a length of one day between 2006 and 2010, approximately 6% of guilty charges and 15% of charges with a custodial sentence. This is nearly identical to the proportion of one-day sentences in the 2002-2006 courts records used to produce the first set of weights. The prevalence of one-day sentences varies depending on the violation. For example, 26.3% of the custody sentences for disturbing the peace (UCR violation 3430) had a length of custody equal to one day, while 1.6% of guilty charges of attempted murder had a length of custody equal to one day.

5.3 Youth data compared to adult data

The Young Criminal Justice Act (UCR violation code 6450) was introduced in April 2003. Since its introduction, the incarceration rate for young offenders has declined for all offences committed by youth. The average length of incarceration has been mainly stable or increased for young offenders since the most serious cases are now the ones that receive a custody sentence. In general, youth have much lower average incarceration lengths and rates than adults. Overall, due to the small proportion of youth per offence and the fact that the CSI relies on *relative* weights, these modifications have not significantly impacted the weights. There were about 406,000 youth charges out of the 3 million charges (less than 15% of the data).

The crime severity weights are produced by combining youth and adult data together. The only time adult and youth data are separated is during the outlier detection step which identifies extremely high lengths of custody, by violation and by adult versus youth.

6. Crime Severity Index

6.1 Most significant violations

Not all violations have the same impact on the CSI. A violation's influence depends on the volume of crime and the magnitude of the weight. In this section the role of the most influential violations to the national CSI are presented. To analyse the impact of the new weights on the CSI, published 2011 data was used. A comparison between results using the original CSI weights (based on 2002-2006 ICCS data) and the revised CSI weights (based on 2006-2010 ICCS data) was done and the main results are available in this section.

5 Most severe violations

The most severe violations changed very little with the introduction of the new weights. The entry of Heroin Import/Export into the top 5 is a result of its weight being changed from a proxy to a calculated weight and our distinguishing it from Heroin Trafficking since 2008.

Table 2: Violations with the highest CSI weights

Revised CSI weights (2006-2010 ICCS data)					
Violation		Rank	Average incarceration (days)	Incarceration rate	CSI weight
1110	Murder 1st Degree	1	7,963.32	94.87%	7,554.94
1120	Murder 2nd Degree*	2			7,554.94
1130	Manslaughter	3	2,070.72	86.04%	1,781.68
4310	Importation and Exportation: Heroin	4	1,864.12	94.44%	1,760.56
1210	Attempted Murder	5	2020.5	85.78%	1,733.14
Original CSI weights (2002-2006 ICCS data)					
Violation		Rank	Average incarceration (days)	Incarceration rate	CSI weight
1110	Murder 1st Degree	1	7,859.10	89.60%	7,041.75
1120	Murder 2nd Degree*	2			7,041.75
1130	Manslaughter	3	2,077.07	87.70%	1,821.56
1210	Attempted Murder	4	1,714.65	82.29%	1,411.01
1520	Hostage Taking	5	1,521.44	84.00%	1,278.01

*1st and 2nd degree murders are combined for the purposes of weight calculation

Top 5 CSI contributors

The largest contributors to the CSI are listed below in table 3, (i.e. those violations with the highest product of volume and weight), have remained relatively unchanged after updating the weights. For the Overall CSI and Non-violent CSI only fraud and mischief have changed places, with all other violations seeing their percentage of contribution changing but leaving their order unchanged. For the Violent CSI all five violations maintain their original order even while some weights saw significant revisions.

Table 3: Top 5 contributors to the CSIs (2011 UCR counts)

Overall CSI						
Violation		UCR Count	Original weight	Revised weight	Contribution to CSI, original weight	Contribution to CSI, revised weight
2120	Breaking & Entering	180,460	186.99	216.67	22.65%	24.52%
1610	Robbery	29,730	583.32	523.33	11.64%	9.76%
2160	Fraud	77,921	108.74	121.85	5.69%	5.95%
2170	Mischief	315,852	29.73	28.21	6.30%	5.59%
2140	Theft $5,000s and Under	221,492	37.41	37.45	5.56%	5.20%
Violent CSI						
Violation		UCR Count	Original weight	Revised weight	Contribution to CSI, original weight	Contribution to CSI, revised weight
1610	Robbery	29,730	583.32	523.33	38.15%	33.24%
1330	Sexual Assault	21,283	210.98	239.84	9.88%	10.91%
1430	Assault, level 1	172,770	23.43	24.53	8.91%	9.05%
1420	Assault, level 2	50,184	77.38	83.36	8.54%	8.94%
1627	Utter Threats to Person	71,945	46.39	45.71	7.34%	7.03%
Non-violent CSI						
Violation		UCR Count	Original weight	Revised weight	Contribution to CSI, original weight	Contribution to CSI, revised weight
2120	Breaking & Entering	180,460	186.99	216.67	32.59%	34.71%
2160	Fraud	77,921	108.74	121.85	8.18%	8.43%
2170	Mischief	315,852	29.73	28.21	9.07%	7.91%
2140	Theft $5,000s and Under	221,492	37.41	37.45	8.00%	7.36%
2142	Theft $5,000s and Under from Motor Vehicle	186,665	37.41	37.45	6.74%	6.21%

6.2 Linkage Factor

As stated earlier, a linkage factor is necessary to maintain year-over-year CSI comparisons after the weights have been revised. The linkage factor ensures that at the national level, for the linkage year, the CSIs (overall, violent, non-violent) are the same whether the original or revised weights are used. Since the linkage factor is only calculated at the national level, CSIs calculated for the linkage year for sub-national regions may change using the new weights. The effect of the new weights will be particularly pronounced in regions where the distribution of crime is different from the national distribution. The table 4 below provides the linkage factor for the three national CSIs, using UCR volumes for 2011:

Table 4: Effect of linkage factor on national CSIs, for linkage year

CSI	Linkage factor	CSI, original weights	CSI, revised weights without linkage factor	CSI, revised weights with linkage factor
Overall	0.934	77.62	83.06	77.62
Violent	0.971	85.26	87.79	85.26
Non-violent	0.919	74.68	81.25	74.68

Recall that the linkage factor is the sum of weighted volumes of crime using the original weights, divided by the sum of weighted volumes of crime using the revised weights (formula 1.3). In other words, using 2011 UCR volumes at the national level, the weighted volumes using the old weights are 93.4% of those obtained using the new weights. In this sense, the linkage factor can be thought of as a deflation factor that preserves the scale of the original weights. The purpose of the weights is to determine the relative, rather than absolute, severity of crime.

Note that this paper uses the 2011 UCR data that was published in 2012. In practice, the linkage factor uses the revised 2011 UCR data (published in 2013). Using the first version of the 2011 data allows us to compare previously published 2011 sub-national CSIs, which were calculated using the original weights, with those that would be obtained using the new, revised weights.

6.3 Ranking Changes

After applying the linkage factors, while most of the sub-national CSIs for 2011 are similar whether the new or old weights are used, the rankings do change slightly at the provincial and territorial, CMA and respondent (Police Service) levels (see table 5 and 6 for more details[7]). Across all CMAs: 3% of CMAs saw their overall CSI change by more than two points with the new weights; 14% saw their violent CSI change by more than two points; and none of the non-violent CSIs at the CMA level changed by more than 2 points. At the Police Service level: 10% saw their overall CSI change by more than 5 points; 32% saw their violent CSI change by more than 5 points; 5% saw their non-violent CSI change by more than 5 points. (Smaller regions see greater fluctuations in their CSIs).

These changes in the regional CSIs are due to the fact that, for some regions, the distribution of crime differs from the national distribution. If, according to the new weights, a region's distribution of crime is less serious than it was using the original weights, then its CSI ranking relative to other regions may drop. If, however, the new weights indicate that the region has more serious crime than it did with the previous weights, then its 2011 CSI ranking relative to other regions may rise. For example, with the new weights, break and enter has a higher weight than before. If a region has proportionally more break and enters than occur nationally, its 2011 CSI ranking using the new weights may be higher than it was with the previous weights.

With the original CSI weights, in 2011, Prince Edward Island had the third lowest overall CSI of all the provinces and territories. Using the new weights, it dropped one rank, having the second lowest overall CSI. At the CMA level, Guelph had the lowest overall CSI in 2011, using the old weights. With the new weights, it still has the lowest overall CSI in 2011.

[7] Table at the respondent level not included in this paper due to the large number of respondents (police services).

Table 5: 2011 CSI rankings by province and territory, original versus new weights (low rank = low CSI)

2011 as Linkage Year	Rank 2011 (published, original weights)			Rank 2011 (new weights)			Rank Change		
Jurisdiction	Overall	Violent	Non-violent	Overall	Violent	Non-violent	Overall	Violent	Non-violent
Canada									
Newfoundland and Labrador	5	2	6	4	2	6	1	0	0
Prince Edward Island	3	1	4	2	1	4	1	0	0
Nova Scotia	6	6	5	6	6	5	0	0	0
New Brunswick	2	3	2	3	3	2	-1	0	0
Quebec	4	5	3	5	5	3	-1	0	0
Ontario	1	4	1	1	4	1	0	0	0
Manitoba	9	10	9	9	10	9	0	0	0
Saskatchewan	10	9	10	10	9	10	0	0	0
Alberta	7	7	7	7	7	7	0	0	0
British Columbia	8	8	8	8	8	8	0	0	0
Yukon	11	11	11	11	11	11	0	0	0
Northwest Territories	13	12	13	13	12	13	0	0	0
Nunavut	12	13	12	12	13	12	0	0	0

Table 6: 2011 CSI rankings by CMA, original versus new weights (low rank = low CSI)

2011 as Linkage Year	Rank 2011 (published)			Rank 2011 (new)			Rank Change		
CMA	Overall	Violent	Non-violent	Overall	Violent	Non-violent	Overall	Violent	Non-violent
St. John's, NFLD	27	19	30	28	20	30	-1	-1	0
Halifax, NS	23	29	21	23	29	20	0	0	1
Moncton, NB	16	13	15	15	16	15	1	-3	0
Saint John, NB	21	25	18	19	25	17	2	0	1
Saguenay, QC	17	8	20	18	11	19	-1	-3	1
Québec, QC	2	2	3	2	1	3	0	1	0
Sherbrooke, QC	8	7	13	7	7	13	1	0	0
Trois-Rivières, QC	15	1	19	16	6	21	-1	-5	-2
Montréal, QC	22	26	17	22	27	18	0	-1	-1
Ottawa-Gatineau, QC	12	12	8	14	18	8	-2	-6	0
Ottawa-Gatineau, ON	4	11	4	4	10	4	0	1	0
Kingston, ON	6	4	12	6	2	11	0	2	1
Peterborough, ON	9	10	9	11	9	10	-2	1	-1
Toronto, ON	3	24	1	3	22	1	0	2	0
Hamilton, ON	13	20	6	12	17	5	1	3	1
St.Catharines-Niagara, ON	7	3	14	8	4	14	-1	-1	0
Kitchener-Cambridge-Waterloo, ON	11	14	5	10	13	6	1	1	-1
Brantford, ON	26	22	28	26	24	26	0	-2	2
Guelph, ON	1	5	2	1	5	2	0	0	0
London, ON	20	15	24	20	14	25	0	1	-1
Windsor, ON	10	9	11	9	8	12	1	1	-1
Barrie, ON	5	6	7	5	3	7	0	3	0
Sudbury, ON	19	21	22	21	21	22	-2	0	0
Thunder Bay, ON	31	31	29	31	31	29	0	0	0
Winnipeg, MN	30	33	23	30	33	23	0	0	0
Regina, SK	33	30	33	33	30	33	0	0	0
Saskatoon, SK	32	32	32	32	32	32	0	0	0
Calgary, AB	14	17	10	13	12	9	1	5	1
Edmonton, AB	25	28	25	24	28	24	1	0	1
Kelowna, BC	29	23	31	29	23	31	0	0	0
Abbotsford-Mission, BC	24	18	27	25	19	28	-1	-1	-1
Vancouver, BC	28	27	26	27	26	27	1	1	-1
Victoria, BC	18	16	16	17	15	16	1	1	0

7. Conclusion

This year was the first time the Crime Severity Index weights were updated since their initial creation and it was an opportunity to refine the methodology used to calculate them. In some instances (drugs and motor vehicle theft), we had additional information that allowed us to create weights that more accurately reflect the delineation seen in the UCR codes. In other instances (outlier detection), we were able to find superior methods which allowed us to keep more valid data in the calculation of the final weights. Overall, the impact of these changes has been small, and, as anticipated, the updating of the weights has had minimal impact on the final CSI values. In another five years, it is anticipated that we will perform a second update of the CSI weights at which time it will be another occasion to review the methodology and take into account in the CSI weights any legislative and sentencing pattern changes and analyze the effects on the CSI values.

8. REFERENCES

Babyak, C. (2008), "Specifications for Calculating the Canadian Crime Severity Index", Statistics Canada internal document.

Babyak, C., Alavi,A., Collins, K., Halladay, A., Tapper, D. (2009), "The Methodology of the Police-Reported Crime Severity Index", Working Paper HSMD-2009-006E/F.

Bernier, J., Nobrega, K. (1998), "Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma gap method" SSC Annual meeting – Proceedings of the survey method section.

Statistics Canada (2003), "Canadian Crime Statistics, 2003", Canadian Centre for Justice Statistics, Statistics Canada, accessible on the internet at http://www.statcan.gc.ca/pub/85-205-x/85-205-x2003000-eng.pdf, accessed July 2009.

Statistics Canada (2008), "The Consumer Price Index - Data quality, concepts and methodology", Statistics Canada, accessible on the internet at http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=2301&lang=en&db=imdb&adm=8&dis=2, accessed July 2009.

Evra, R., Franklin, S., Babyak, C., Campbell, A. (2013), Statistics Canada internal "Updating the Police-Reported Crime Severity Index Weights".