Revised CIS-2008 weighting procedure description

Nico Trocmé, McGill University
Vandna Sinha, McGill University
Barbara Fallon, University of Toronto
Bruce MacLaurin, University of Calgary

Posted on www.cwrp.ca on January 25, 2012

Centre for Research on Children and Families
McGill University, 3506 University Street, Suite 106
Montréal, Québec, Canada H3A 2A7
Tel: (514) 398-5286   Fax: (514) 398-5287
http://www.mcgill.ca/crcf/

Revised CIS-2008 weighting procedure description

The data collected for the CIS 2008 were weighted in order to derive national annual incidence estimates, first by applying a composite regionalization weight and then by applying an annualization weight. This document describes the procedure used to derive these weights

The overall weights used to derive national annual estimates described in this revised weighting document are the same as the overall weights described in previous CIS reports including the CIS-2008 Major Finding report[1]. The revised wording differs from the original in two ways: (1) the original annualization weight combined the annualization weight and the subsampling weight, the subsampling weight is now included as a component of the revised regionalization weight, and (2) the original regionalization weight did not separate out the sample weight from the agency size correction, the revised procedure identifies both components separately. While mathematically there has been no change to the final weight applied to each site, the revised wording provides a more accurate statistical interpretation of the weights used to derive national annual estimates.

To simplify the terminology in this document the term "agency" is used to refer to the primary sampling unit in the study, the local organization responsible for conducting child-maltreatment related investigations. In some jurisdictions, these organizations were autonomous agencies; in others, they were local offices for the provincial or territorial child protection authority.

This document is organized in two sections: the first provides a summary description of the weighting procedures including limitations of the assumptions underlying the weights, the second is a more detailed technical description of the procedure.

Summary Description

The data collected for the CIS 2008 were weighted to derive national annual incidence estimates by applying a composite regionalization weight and an annualization weight. The regionalization weight was developed to estimate the number of investigations completed within the three-month data collection period by child welfare organizations across Canada. The regionalization weight includes three components: (1) a sample weight that adjusts for the disproportional selection of agencies from oversampling provinces, (2) a subsampling weight that accounts for random subsampling of investigations in agencies that investigated more than 250 cases during the three-month data collection period, and (3) an agency size correction, designed to adjust for variations in the size of agencies within a stratum. The annualization weight is used to estimate annual investigation volume based on the investigation volume during the three month data collection period of CIS-2008. The annualization weight is the ratio of all investigations conducted by a sampled agency during 2008 to investigations conducted by the sampled agency during the case selection period.

Three limitations to this estimation method should be noted. The agency size correction uses child population as a proxy for agency size; this does not account for variations in per capita investigation rates across agencies in the same strata. The annualization weight corrects for seasonal fluctuation in the volume of investigations, but it does not correct for seasonal variations in types of investigations conducted. Finally, the annualization weight includes cases that were investigated more than once in the year as a result of the case being re-opened following a first investigation completed earlier in the same year. Accordingly, the weighted annual represent the child maltreatment-related investigations, rather than investigated children.

---

[1] Trocmé, N., Fallon, B., MacLaurin, B., Sinha, V., et al. (2010) *Methodology* in Canadian Incidence Study of Reported Child Abuse and Neglect – 2008: Major Findings. Public Health Agency of Canada: Ottawa, 2010.

Detailed Technical Description

The data collected for the CIS 2008 were weighted to derive national annual incidence estimates by applying a composite regionalization weight and an annualization weight. The regionalization weight was developed to estimate the number of investigations completed within the three-month data collection period by child welfare organizations across Canada. The annualization weight is used to estimate annual investigation volume based on the investigation volume during the three month data collection period of CIS-2008.

Regionalization weight

The regionalization weights were used to account for the sampling and subsampling used during the three month data collection period. The weight is composed of a sample weight, a subsampling weight, and an agency size correction.

Sample weight – The CIS‑2008 sampled a higher proportion of agencies in British Columbia, Alberta, Saskatchewan, Ontario and Quebec. These five provinces supported inclusion of a sufficient number of agencies in the CIS‑2008 sample in order to enable analysis of province-specific data. As a result, the proportion of agencies sampled in these provinces was higher than the proportion sampled in other provinces/territories, and the unweighted data disproportionately reflects the investigation rates and profiles of the oversampling provinces. The first factor, the "sample weight" or "W_s", adjusts for the disproportional selection of agencies from oversampling provinces. This weighting factor represents the ratio of the total number of agencies in a stratum (a group of child welfare organizations within a geographic region from which sites were randomly sampled) to the number of agencies sampled from that stratum. For example if we sampled only one agency from a province with 10 agencies, that agency would have been given a weight of 10 (10/1). In contrast, if we consider the example of an oversampling province with 10 agencies, from which we sampled 4 agencies, each of the sampled agencies would be assigned a weight of 2.5 (10/4)

It should be noted that some sites were not randomly sampled, either because they represented a large metropolitan centre that was automatically included in the study, or were from Quebec or Saskatchewan where all provincial agencies were included in the study. In these instances the sample weight was 1.

$$W_s = \frac{\#\ of\ agencies\ in\ stratum}{\#\ of\ agencies\ sampled\ in\ stratum}$$

Subsampling weight – In most agencies, data were collected for every new, maltreatment-related investigation opened during the three month data collection period; however, in order to reduce burden on workers, sample size was limited to 250, randomly selected investigations in 20 very large agencies and data on one out of two investigations was selected for data collection in 16 Quebec agencies. The subsampling weight – W_{ss} – accounts for this random subsampling of investigations within the three-month data collection period. This factor represents the ratio of the number of investigations opened by an agency during the three-month data collection period to the number of investigations from that agency which were included in the CIS sample. For example, a subsampling weight of 4 (1,000/250) would have been assigned to cases from an agency where data were collected for a random sample of 250 cases in an agency that investigated 1,000 cases during the data collection period.

$$W_{ss} = \frac{\#\ of\ investigations\ Oct\ 1 - Dec\ 31}{\#\ of\ investigations\ sampled}$$

Agency Size Correction – Child welfare organizations, including those in the study sample, vary greatly in terms of the number of children they serve and the number of investigations they conduct. The "sample weight" described above adjusts for differences in the number of agencies selected from each stratum, but does not account for variations in the size of the agencies within these strata. The third component of the regionalization weight, which we can call PS_r, is designed to adjust for variations in the size of agencies within a stratum. It represents the ratio of the average child population for all the agencies in the stratum to the average child population served by the agencies sampled within that stratum. For example in a stratum of 10 agencies serving 100,000 children (average

child population in agencies in stratum = 10,000), one sampled agency serves a region with 6,000 children and the second serving a region with 18,000 children (average child population in sampled agencies = 12,000), the agency size correction would be 10,000 / 12,000 = 0.83.

$$PS_r = \frac{\text{average child population in stratum}}{\text{average child population in sampled agencies}}$$

An important limitation to the method used to derive the agency size correction must be noted. Ideally, this factor would adjust for variations in the number of investigations opened by agencies within a stratum. But, because reliable statistics on number of investigations completed by an agency have not been consistently available, child population is used as a proxy for agency size[2]. Accordingly, this factor assumes that the numbers of investigations opened by the agencies within a stratum are strictly proportional to agency child population and it does not account for variations in the per capita rate of investigations.

Regionalization Weight: Together, these three factors, $W_s \times W_{ss} \times PS_r$, are used to create the regionalization weights which are used to estimate the number of investigations completed within the three-month data collection period by all child welfare organizations in Canada. This would mean, using the examples provided above, that every sampled case from the oversampling province agency illustration would receive a weight of $2.5 \times 4 \times 0.83 = 8.33$.

Annualization weight

Because the CIS collects data only during a three-month period from sampled child welfare agencies, data from the agencies were weighted to estimate the number of investigations conducted by the sampled agencies during the full year. Accordingly, all data were multiplied by an annualization weight, which we can call PS_a, which represents the ratio of all investigations conducted by a sampled agency during 2008 to all investigations opened by the sampled agency during the case selection period: Oct 1 – Dec 31 2008. For example, if an agency conducted 1,800 investigations during in 2008, 500 of which were investigated from October 1 to December 31, the annualization weight would be $1,800/500 = 3.6$.

$$PS_a = \frac{\#\ of\ investigations\ in\ 2008}{\#\ of\ investigations\ oct\ 1 - dec\ 31}$$

Two key limitations of the annualization weights must be noted. This factor corrects for seasonal fluctuation in the number of investigations, but it does not correct for any seasonal variations in the types of investigations conducted. In addition, while cases reported more than once during the three-month case sampling period were unduplicated (see *Case Selection* section in this chapter), the weights used for CIS-2008 annual estimates include cases that were investigated more than once in the year as a result of the case being re-opened following a first investigation completed earlier in the same year. Accordingly, the weighted annual represent new child maltreatment-related investigations conducted by the sampled agencies in 2008, rather than investigated children.

Full weight (W_{RA})

The weight used to derive national annual estimates, called W_{RA}, is the agency specific weight that is the product of the regionalization weight by the annualization weight. Using the examples developed above, cases from the oversampling agency illustration would be given a final weight of 3.6 x 8.3 = 30.

$$W_{RA} = Ws \times Wss \times PS_r \times PS_a$$

---

[2] This approach was originally developed for the 1993 OIS and used in the 1998 CIS, which built on OIS methods, because, at the time, most jurisdictions could not report on investigation counts and there were dramatic discrepancies in the counts reported. While the quality of investigation statistics has improved, we continue to find important discrepancies in the ways investigations statistics are reported. Site researchers carefully review all case counts provided by the child welfare authorities participating in the study, however, this level of quality control is not available for authorities that were not part of the CIS sample.