

Resolving Differences in Statistical Units: Statistics Canada's Experiences with Using Administrative Data in Economic Programs

Leon Jang¹

Abstract

Administrative data are used more and more in Statistics Canada's economic programs. For the administrative data to become usable for the production of economic statistics, several processing steps are needed. For instance, edit and verification procedures are established to ensure accuracy and consistency in the administrative records, which are also often used in the stratification of survey units, the sample allocation and the detection of outliers. As well, to produce the final estimates, the survey data are calibrated to ensure that the survey results respect totals from the administrative sources.

This paper explains the challenges in integrating administrative data throughout Statistics Canada's economic programs and the methods used to address problems encountered by their use. This will be illustrated through the survey process of the Canadian Survey of Employment, Payrolls and Hours (SEPH). SEPH is a monthly survey conducted by Statistics Canada to produce estimates of levels and trends in the number of employees, earnings and hours worked by detailed industry and geography. The survey design uses two sources of data: a survey of 15,000 establishments and a census of payroll deductions coming from administrative records. Administrative information for total gross monthly payrolls and the total number of employees are obtained from payroll deduction accounts maintained by the federal revenue agency. As this data is provided at the form level, it must eventually be transformed to the establishment level required for the survey. With that goal, processing steps are carried out at intermediate statistical unit levels in order to ensure that aggregations are interpretable and contain minimal discrepancies. Additional processing of the different reporting periods of the administrative payroll data must also respect SEPH's monthly concept in order to respect data coherence.

Key Words: Administrative data, business survey, processing

1. Introduction

As authorized by the Statistics Act (1985), Statistics Canada has the mandate to collect and obtain information to be used for statistical or analytical purposes. Included in this mandate is access to administrative data from businesses, governments or other organizations. Among the strategies described in the agency's Corporate Business Plan (Statistics Canada 2015) are to increase the use of relatively inexpensive administrative data sources to complement more expensive survey data and to increase the efficiency in administrative operations in order to direct more resources to statistical programs.

¹ Leon Jang, Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, ON K1A 0T6 Canada, email: leon.jang@canada.ca.

For many years, Statistics Canada has obtained tax data from the Canada Revenue Agency (CRA), the federal agency responsible for administering tax laws for the Government of Canada. Within Statistics Canada, the Administrative Data Division is charged with acquiring tax data from CRA and other administrative data sources that are required by different statistical programs and providing support in dealing with methods or processes associated with administrative data.

Among the surveys using administrative data is the Survey of Employment, Payrolls and Hours (SEPH). Monthly payroll employment data have been collected by the Government of Canada since 1918, with Statistics Canada (then known as the Dominion Bureau of Statistics) taking responsibility for payroll statistics in 1922. Data for earnings were collected starting in 1941 and with the addition of earnings and hours data in 1983, SEPH was created with a sample size of about 70,000 establishments (Schiopu-Kratina and Srinath 1986). With the aim of reducing response burden associated with the survey, CRA made an agreement with Statistics Canada to add two fields on employment and gross monthly payroll information to its payroll deductions remittance (PD7) forms (Rancourt and Hidioglou 1998). These fields were first added to PD7 forms in 1993, allowing for a reduction in sample size to 29,000 units (Dolson 1994) and later to 18,000 units (Hidioglou *et al.* 1995). The two administrative data fields continue to enhance the estimates produced by SEPH to this date, with the current sample now consisting of 15,000 establishments.

While the use of administrative data has enabled SEPH to reduce survey costs and response burden, its integration into the survey is not seamless. Much processing is required for the information received from PD7 forms. This work includes validating form fields, eliminating duplicates and performing edit and imputation. It is only after the processing of the PD7 forms is complete that the data can be aggregated to the statistical levels required for subsequent steps.

In section 2, an overview of SEPH is given. Section 3 presents the payroll deduction administrative data used in SEPH. The processing of the administrative data, including aggregation of the data to subsequent statistical levels, is described in section 4. In section 5, the issues and challenges related to the potential impact of processing are outlined. Finally, in section 6 some concluding remarks are provided.

2. The Survey of Employment, Payrolls and Hours

SEPH is Statistics Canada's monthly data source of detailed information on the total number of employees, payrolls and hours at detailed industry and geography levels. The survey provides the principal input to labour income estimates and also serves as a proxy output measure for approximately 15% of real gross domestic product. SEPH data are also used in the calculation of equalization payments to provinces by the federal government, in the revision of maximum pensionable earnings and retirement savings plan contribution limits by CRA and in contract escalation and wage rate determination by the private sector.

The target population for SEPH is composed of all businesses in Canada having at least one employee and, thus, issued at least one payroll deduction remittance during the reference month. Excluded are businesses primarily involved in agriculture, fishing, hunting and trapping, private household services, religious organizations, international and other extraterritorial public administration and military personnel of defence services.

Industrial levels are based on the North American Industry Classification System (NAICS). See Statistics Canada (2012) for further details on this industry classification.

Survey estimates for SEPH are produced on a monthly basis using a combination of three different data sources: a census of payroll deductions, provided by CRA; the Business Payrolls Survey (BPS), which collects data from a sample of 15,000 businesses; and payroll file extractions for different levels of government.

2.1 Administrative Records

CRA provides to Statistics Canada a file of administrative data from payroll deduction remittance forms (PD7). Two variables of interest for SEPH are found on the PD7 forms: the total gross payrolls and the total number of employees for the reporting period. In addition, the amount of income tax remitted to CRA by the employer is also available from the PD7, along with Employment Insurance premiums and either Quebec Pension Plan contributions for employers in the province of Quebec or Canada Pension Plan contributions for employers from other provinces or territories. The remittance amount, which includes contributions from both the employer and employees, is highly correlated with gross payrolls. Further details on the administrative records received and the PD7 forms utilized to collect the data are discussed in section 3.

2.2 The Business Payrolls Survey

The monthly BPS is the questionnaire component of SEPH. It is used to estimate variables for SEPH that are not available from administrative data, such as information on weekly payroll, number of regular and overtime hours worked and special payments, with all these variables disaggregated by employment category (i.e., employees paid by the hour, salaried employees, other employees). The BPS also collects information on job vacancies. Completion of the BPS is a legal requirement under the Statistics Act.

A SEPH target population of around 1,000,000 establishments is created from Statistics Canada's Business Register, a centralized frame for all Canadian businesses (Statistics Canada 2010). From this population, a sample of around 15,000 establishments is selected, stratified based on province or territory, NAICS subsector (three-digit code) or industry group (four-digit code) and establishment size (number of employees). Every month, the Business Register is updated using survey feedback, business profiling and administrative data, and one-twelfth of the sample is rotated out and replaced. However, around 800 establishments are permanently in the sample due to a significant share of employment or payroll or because they have payroll characteristics that are different from other establishments within their strata.

2.3 Government Files

Most of the information about general government services at the provincial, territorial and federal levels is provided to Statistics Canada on a monthly basis in the form of electronic files extracted from their payroll lists. A special BPS – Public Sector questionnaire is used to collect data for government entities that are not included in the electronic files.

3. Payroll Deductions

As mentioned in section 2.1, Statistics Canada receives a census of PD7 forms from CRA. Canadian employers are required by law to remit to CRA Canada Pension Plan and Quebec Pension Plan contributions, Employment Insurance premiums and income tax deducted

from employees' income along with the employer's own share of contributions and premiums. Employers are required to apply for a business number (BN) and register for a payroll program account with CRA in order to be a remitter (Canada Revenue Agency 2016b).

3.1 Remitter Types

The frequency of payroll deduction remittance is determined on two factors: the average monthly withholding amount (remittance) and perfect compliance history. Based on these factors, an employer can become eligible for three types of remitters. Each type of remitter must remit deductions on a different basis. See Canada Revenue Agency (2016a) for complete details on remitting payroll deductions.

3.1.1 Regular remitter

Any employer with an average monthly remittance of less than \$25,000 CAD (Canadian dollars) is a regular remitter. New employers are also classified as regular remitters. Regular remitters must remit their deductions to CRA by the 15th day of the month following the month when the deductions were made.

3.1.2 Quarterly remitter

Employers having an average monthly remittance of less than \$3,000 CAD in either of the previous two calendar years become eligible to remit deductions on a quarterly basis provided they have perfect compliance history. In order to achieve perfect compliance, an employer must have made all remittances on time, paid their collected sales tax on time and have filed their annual income tax returns on time, all over a 12-month period.

In addition, new employers with monthly remittances of less than \$1,000 CAD can also become quarterly remitters. However, these new employers revert to being a regular remitter once the remittance for any month exceeds the \$1,000 CAD threshold.

Quarterly remitters must remit their deductions to CRA on January 15, April 15, July 15 and October 15 for the preceding three-month period. For example, deductions made from April to June are due by July 15.

3.1.3 Accelerated remitter

When an employer has an average monthly remittance greater than or equal to \$25,000 CAD, it is considered to be an accelerated remitter. Multiple remittances are expected to be sent to CRA during a month. Note that employers will not necessarily remit more frequently than their payroll frequency (e.g., employees paid monthly, twice monthly or biweekly) as the date of each deduction determines when it is due to CRA.

Accelerated remitters with an average monthly remittance less than \$100,000 CAD must remit their deductions to CRA twice per month. For remuneration paid in the first 15 days of the month, deductions are due by the 25th day of the same month. For remuneration paid from the 16th day to the end of the month, deductions are due to CRA by the 10th day of the following month.

Employers with an average monthly remittance of \$100,000 CAD or more are required to remit their deductions four times per month. CRA divides each month into four periods: the 1st to the 7th day of the month, the 8th to the 14th, the 15th to the 21st and the 22nd to the end of the month. When remunerations fall within any period, deductions are due to

CRA by the third working day (excluding weekends and public holidays) following the end of the specified period.

3.2 Remittance Forms

There are two general types of PD7 forms for remittances: PD7A (regular) and PD7A(TM) (accelerated). The regular PD7 form is used by regular and quarterly remitters, while the accelerated PD7 form is used by accelerated remitters that have multiple remitting periods per month. Of the 1,000,000 businesses in-scope for SEPH, about 800,000 remit using the regular form, representing 20% of employment. The remaining 200,000 businesses remit using the accelerated form, and these businesses represent 80% of all employment.

The PD7 forms contain information filled out by employers that is of interest to SEPH and received from CRA. Information that are common to regular and accelerated forms are:

- Business number (BN)
- Gross payroll for the remitting period (rounded to the nearest dollar)
- Number of employees in the last pay period (of the remitting period)
- End of the remitting period
- Amount paid (remittance) for the remitting period

For the regular PD7 form, the end of the remitting period is the year and month for regular remitters or the last year and month of the quarter for quarterly remitters. On the accelerated PD7 form, the end date of the remitting period (year, month and day) is given. Larger businesses can have multiple BN accounts and fill many forms during a month.

4. Processing of the Administrative Data

Statistics Canada does not share the same data needs as CRA for the payroll deduction information described in section 3. While the main interest of CRA is to collect money owed in the form of remittances, SEPH is instead interested in collecting data on payroll and employment. Also, the BNs used by CRA do not align with the statistical structure of Statistics Canada that requires data at the establishment level. Fortunately, the two structures can be linked using a common identifier as the business number prefix (BN root) is available in Statistics Canada's Business Register. This BN root provides the link between processed payroll deduction data and establishment-level data. This linkage is shown in Figure 1.

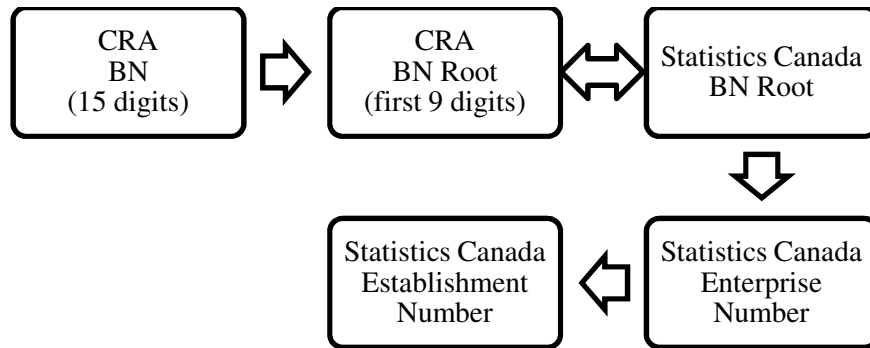


Figure 1: Linkage between Canada Revenue Agency and Statistics Canada identifiers

Several steps are required in order to transform the payroll deduction administrative data at the PD7 form level prior to producing establishment-level data. This includes PD7 form edits, the aggregation of multiple forms by PD7 account, imputation at the BN level, validation of aggregated data and allocation to the establishment level. All this processing is done in order to meet Statistics Canada's dimensions of quality (Statistics Canada 2009), particularly in terms of coherence with respect to common concepts and of accuracy in ensuring identical coverage between data sources.

4.1 Form Edits

The ultimate goal of PD7 forms is to collect deduction remittances from employers. As such, an internal validation check is done by CRA to compare the remittances reported on the forms with the actual dollar amounts remitted by employers. No such validation is done by CRA for employment and gross payroll values, and it is possible for a PD7 form to be submitted without these values. Consequently, remittance values are more likely to be verified by CRA than the two fields utilized by SEPH (Yeung *et al.* 2008).

A series of PD7 form-level edit rules has been established to address SEPH's use of the administrative data. Most of the edits focus on editing the payroll variable. At the form level, it is more difficult to identify atypical values. Employment differences are therefore not verified until aggregation to the BN, which is explained in section 4.2.

4.1.1 Form verifications

The Government of Canada's business number initiative allows for businesses to have a single BN for four major types of accounts: customs, import and export (RM); payroll source deductions (RP); corporate income tax (RC); and Goods and Services Tax (RT). PD7 forms should only relate to RP business numbers. SEPH verifies that all PD7 forms received are for RP accounts. When they are not, the forms are removed from further processing.

In addition, duplicate PD7 forms are sometimes transmitted to Statistics Canada. When this occurs, the edit system identifies and removes the duplicates.

Finally, the system also identifies PD7 forms that contain erroneous employment or payroll values at the form level. Validity rules are applied, and extreme values are identified based on ranges that are considered unusual. These special cases are later resolved at the BN level as explained in section 4.2.

4.1.2 Validity edits

Transcription errors sometimes occur when employers fill out the PD7 forms. Deterministic imputation is performed for two possible errors. When the end of the remitting period is left blank, which is often the case for employers using the regular PD7 form, the value for the current reference month is assigned to the form. Also, the gross payroll for the remitting period is adjusted when the value is clearly entered in cents rather than dollars. In this case, the payroll value is divided by 100.

4.2 Aggregation and Imputation of Multiple Forms

To obtain a monthly value, the PD7 form data must be summarized for each business. Each form represents a single deduction remittance, as described in section 3.2, and many remittance periods exist within a month for accelerated remitters. Therefore, a complex set

of aggregation rules is used to determine the appropriate employment values. Imputation for missing information also occurs at this stage.

A PD7 form provides the number of employees in a given last pay period, so a monthly value cannot be calculated by cumulating multiple forms as each one can refer to the same group of employees. A method to determine the correct employment for the month that meets SEPH's concept needs to be done. Meanwhile, the gross payroll for each remitting period can be cumulated to an equivalent monthly value. However, this will only be correct if all forms are received and all forms contain valid payroll values.

For the purposes of aggregation, SEPH assigns a reference week corresponding to the end date of the remitting period. While mainly used for accelerated remitters, these reference periods are also assigned to regular remitters. Reference weeks refer to end dates from the 1st to 7th (1), 8th to 14th (2), 15th to 21st (3), 22nd to 28th (4) and 29th to the month end (5).

Rules for aggregating forms differ depending on the remitter types described in section 3.1. Each type of remitter applies a different set of rules to determine the monthly employment and gross payroll.

4.2.1 Rules for regular remitters

Only one regular PD7 form is expected each month from each regular remitter. When one complete form is received, the employment and payroll values are retained for the business. However, there are two other possible scenarios that need to be addressed:

1. If no form is received for a regular remitter or if an incomplete form is received without both the employment and payroll values, those missing values are imputed. Imputation methods depend on the information available for the business.
 - If the business appears to be active in the current month and if data from the previous month are of good quality then trend imputation will be done.
 - If data are not available for the previous month but other variables from the current month are available (e.g., payroll is given but employment is not) then ratio imputation will be done.
 - If no information is available for the current or previous months then stratum-level averages for the current month will be used for mean imputation.
2. If more than one complete form is received for a regular remitter, the monthly employment is determined by the maximum employment in any week. When multiple forms are received for the same week, rules exist to determine whether or not to add up employment. Note that exact duplicate forms have already been removed as explained in section 4.1.1. Monthly payroll is added up from all complete forms received for each business.

4.2.2 Rules for quarterly remitters

While quarterly remitters are eligible to remit deductions every three months, most employers actually remit on a monthly basis. In fact, fewer than 20% of eligible quarterly remitters remit their deductions on a quarterly basis. For this reason, SEPH derives a flag to identify true quarterly remitters.

When a quarterly remitter identified by SEPH reports payroll deductions (using the regular PD7 form as explained in section 3.2), the employment value is carried forward while one-third of the payroll value is assigned for the given month. The imputation rules for regular remitters as explained in section 4.2.1 apply when no form is received for the month. This results in a lag as the values for the reporting month become the basis for imputation over the next two months, for which the true values will only be known when the next form is received three months later.

4.2.3 Rules for accelerated remitters

The monthly activity code represents the expected number of weeks or remitting periods for which remittances have been received in the current month or were received in past months. This code is used to determine the expected number of weeks where employment and payroll will be positive (i.e., PD7 forms received). The monthly activity code is equal to the largest number of transactions received in a month over the past four months, up to a maximum value of 4.

Should the number of weeks with forms be fewer than specified by the monthly activity code, the employment is flagged for imputation while the payroll is adjusted by a ratio. For example, when four weeks of data from PD7 forms are expected and only three weeks are received then the monthly payroll is corrected by a factor of 4/3.

Imputation for missing PD7 forms follow the same rules for regular remitters as explained in section 4.2.1. However, the determination of the monthly employment value is different for accelerated remitters dependent on payroll type (i.e., pay frequency). An appropriate monthly value is selected using a set of rules applied according to whether or not employment and payroll are positive during each of the five weeks of a month. In general, the value for Week 4 (22nd to 28th of the month) is used as the base week if employment is positive. Otherwise, the first week less than Week 4 with positive employment is used as the base week. Accelerated remitters with four remitting periods per month use the base week employment for its monthly employment, while accelerated remitters with two remitting periods per month use the sum of the base week and the previous week (usually Weeks 3 and 4) for its monthly employment in order to cover the possible occurrence of multiple forms. This is done in case the remittances refer to different groups of employees.

4.3 Manual Review

Once all PD7 forms have been aggregated by BN, an assessment of significant month-to-month changes occurs, identifying businesses having large shifts in employment or payroll.

Each month, a list of businesses having large changes is generated and sent to survey analysts. For observations that exist in both the current and previous months, a score function is used to prioritize the businesses to review. This score is based on changes in employment, payroll and the ratio of payroll to employment between the two months.

Using a processing environment, analysts can review monthly values and PD7 form data for each business flagged for manual review. Survey analysts can resolve the employment or payroll values.

4.4 Allocation to Establishments

The 15-digit BN for each employer consists of a 9-digit prefix (BN root), a 2-character account type (see section 4.1.1) and a 4-digit sequential account number. With all

employment and payroll values now available for each business, the next step is to aggregate all business numbers by BN root in order to account for businesses with multiple accounts. As shown in Figure 1, the BN root is the key link in order to get the statistical structure used by SEPH.

The Business Register at Statistics Canada is maintained using files of legal entities produced by CRA, which includes information such as the legal name, business address and the main industrial activity of the business. Other information integrated from CRA are business number data (BN root), payroll deduction data and income tax data. For each business, the Business Register derives the four following statistical attributes (Statistics Canada 2010):

- Enterprise, the legal unit or business in most cases
- Company, the lowest level where operating profit can be calculated
- Establishment, the level providing data on profit, salary and wages
- Location, the level providing employment and/or revenue data

SEPH employment and payroll information at the BN root is linked to the Business Register to get the associated enterprise identification number. Employment and pay values are aggregated by enterprise and then allocated down to the establishment level. Data at the establishment level provides the ability to produce SEPH estimates by province and industrial level.

5. Challenges in Administrative Data Processing

Given the nature of combining form-level administrative data with the construct of an establishment-level survey, issues during the data processing steps can occur that can potentially impact the quality of the data. This section addresses the uncertainty presented during different steps covered in section 4 and presents work currently being done to reduce its impact.

5.1 Conceptual Issues

As detailed in section 3.2, the PD7 forms refer to a remittance period. For SEPH, this needs to be associated with a specific reference week (see section 4.2), which in turn is represented within a reference month. When CRA provides the weekly transmission of PD7 data, many reference periods are covered due to the different possible types of remitters. As well, more than one month of weekly files are needed to cover an entire reference month for SEPH.

It is for this reason that SEPH requires two separate months to publish estimates for a reference month. For example, remittances for the January reference period will be received starting in January but also in February and even in later months. SEPH first publishes January estimates in March and revises these estimates in April based on further reception of administrative data from CRA for that reference period.

The main challenge with collecting employment and payroll information for SEPH using the PD7 forms is that the main field for CRA from these forms is deduction remittance. As explained in section 4.1, employment and payroll are not mandatory fields and are not subject to data validation by CRA. The use of an administrative data source is beneficial in providing SEPH with a census for two variables covering all provincial and industrial

levels. However, appropriate methods must be used to account for missing or erroneous values. Established edit, imputation and manual review help to ensure that the data received are relevant, accurate and coherent.

5.2 Aggregation Issues

In section 4.2.1, the method to treat regular remitters having multiple PD7 forms for the same week is discussed. When the forms are not considered to be duplicates, it needs to be determined whether or not to add up employment. At present, if the number of employees is equal or close to the value from other forms then the larger value is retained. Otherwise, the employment values are added up for the week. It is hard to know if the same employees are being covered by multiple forms. Without having any way to identify which employees are represented on a remittance form, it is difficult to tell if there are multiple groups of employees and therefore hard to determine if this method is accurate.

Given that aggregation is done differently for regular and accelerated remitters, this could impact the SEPH estimates produced for different industrial levels. This could be particularly true for industries with more complex businesses that require multiple PD7 forms, which are more prone to edit, imputation and aggregation. By dividing the reference month into five reference weeks, issues in terms of duplicate employment values are addressed.

Businesses with multiple payroll source deduction accounts could possibly remit under any of their accounts. For these situations, the employment reported for these reports could be duplicated when aggregating to the BN root. Historical validations at the business level from month to month can identify changes that require review.

5.3 Imputation Issues

The different methods of imputing employment and payroll presented in section 4.2.1 provide a consistent framework to resolving missing or erroneous values. However, some problems can still arise following imputation that present issues with respect to data coherence and interpretability.

It may seem logical that payroll be non-zero when remittances are greater than zero. Similarly, when there are no remittances for a month it is likely that there should not be any associated payroll for that period. However, PD7 remittances are only used in the form validation step to identify unusual remittance to pay and remittance to employment ratios as explained in section 4.1.1. They are not used in the imputation rules, so it is possible for imputed payroll to be positive when the remittance is zero and for it to be zero when the remittance is positive. Currently, around 2% of businesses are in this situation where the remittance and payroll tendencies do not match (Xing *et al.* 2016).

Provided that the data in the previous month are of good quality, they will be used for trend imputation for the current month when required. This can cause a problem for businesses that are inactive. Should this inactivity not be detected, the values for the PD7 forms that are not received will continue to be imputed. Again, remittance values could be used to validate the expectation of payroll and employment for a given month.

Accelerated remitters, as explained in section 4.2.3, use a monthly activity code that implies the number of weeks or remitting periods for which remittances, employment and pay should be reported. When imputation for payroll is required, ratio imputation is

performed to account for missing periods. However, care must be taken when extra pay periods occur in a month (e.g., employees paid biweekly can receive pay two or three times each month) that could be missed when using this type of imputation.

5.4 Error Detection

As explained in section 4.1.1, employment or pay from PD7 forms can be considered erroneous based on ranges that consider a value to be extreme or unusual. When the employment or payroll values are determined to be erroneous, the processing system will identify these records and correct them later. Around 1% of all businesses are detected as errors using these methods.

However, it has been shown that out of all records detected as erroneous, just 27% receive treatment to correct the extreme values or ratios. Among the remaining errors, 57% are regular remitters that are not treated while 15% are accelerated remitters that are untreated (Xing *et al.* 2016). In this case, the processing system should be revisited so that the error detection and imputation systems properly identify and resolve those records that require correction.

5.5 Current Studies

In light of a survey redesign, different SEPH production issues are being investigated. With respect to the processing challenges described in this section, many studies have begun to address the issues with data coherence seen in processing PD7 form data. The use of the remittance to validate the accuracy and reliability of employment and payroll is being studied in making outlier detection and imputation decisions. Also, a limit on the number of consecutive months a business can undergo trend imputation using data from the previous month is being considered. Finally, the selection of a weekly maximum employment for accelerated remitters rather than the value for one specific week is being investigated.

Annual income tax statements of earnings from employers could be used as another administrative source to verify the consistency of form data. A study performed by Yeung *et al.* (2008) showed that while these annual statements and aggregated annual PD7 data were not equal, identifying known reasons for the differences could significantly reduce or eliminate the discrepancy between the two sources. Additional studies could be completed to see if these tax statements could contribute as a benchmark data source.

6. Concluding Remarks

This paper illustrates the methods used at Statistics Canada to incorporate a census of administrative data with the survey component of the SEPH program. From deduction remittance forms collecting three auxiliary variables (i.e., payrolls, employment and remittances), several steps have been established to validate information and use data reported in different remittance periods to fit survey concepts. Finally, the employment and payroll values for each business are then allocated to Statistics Canada's statistical levels in order to produce estimates at appropriate detailed levels. Despite using statistical methods to maintain the relevance and coherence between data sources, there remain certain processing steps where the quality of the data could be affected. A thorough evaluation of alternate conditions for the processing of administrative data is currently being investigated. While it is difficult to quantify the impact of uncertainty within the data, such work will benefit the future improvement and redesign of SEPH. In spite of these

challenges, the processing of the CRA data done by Statistics Canada ensures that the administrative data being integrated with collected SEPH survey data will be of high quality.

Acknowledgements

The author would like to thank Benjamin Lee, Brett Singbeil and Wesley Yung for their insightful comments. He would also like to thank Danielle Lebrasseur and Geoffrey Logan for their valuable contributions and continued support.

References

- Canada Revenue Agency (2016a). *Employers' Guide – Payroll Deductions and Remittances*. Guide T4001(E) Rev.15.
- Canada Revenue Agency (2016b). *The Business Number and Your Canada Revenue Agency Program Accounts*. Booklet RC2(E) Rev.15.
- Dolson, D. (1994). "On Using a Very Current Source of Frame Information in Canada's Establishment Based Employment Survey." *First Eurostat Workshop on Techniques of Enterprise Panels*. Luxembourg.
- Hidiroglou, M., Latouche, M., Armstrong, B. and Gossen, M. (1995). "Improving Survey Information Using Administrative Records: The Case of the Canadian Employment Survey." *Proceedings of the 1995 Annual Research Conference*. U.S. Bureau of the Census, 171-197.
- Rancourt, E. and Hidiroglou, M. (1998). "Use of Administrative Records in the Canadian Survey of Employment, Payrolls and Hours." *Proceedings of the Survey Methods Section*. Statistical Society of Canada Annual Meeting, June 1998, 39-47.
- Schiopu-Kratina, I. and Srinath, K.P. (1986). *The methodology of the Survey of Employment, Payroll and Hours*. Statistics Canada internal document.
- Statistics Act, Revised Statutes of Canada (1985, c. S-19). Retrieved from the Government of Canada Justice Laws website: <http://laws-lois.justice.gc.ca/eng/acts/S-19/FullText.html>
- Statistics Canada. (2009). *Statistics Canada Quality Guidelines, Fifth Edition – October 2009*. Statistics Canada catalogue number 12-539-X.
- Statistics Canada. (2010). *A Brief Guide to the BR – July 2010*. Statistics Canada, Ottawa, Canada. Retrieved from the Statistics Canada website: http://www23.statcan.gc.ca/imdb-bmdi/document/1105_D2_T1_V3-eng.pdf
- Statistics Canada. (2012). *North American Industry Classification System (NAICS) Canada, 2012*. Statistics Canada catalogue number 12-501-X.
- Statistics Canada. (2015). *Corporate Business Plan, Statistics Canada, 2015-16 to 2017-18*. Statistics Canada, Ottawa, Canada. Retrieved from the Statistics Canada website: <http://www.statcan.gc.ca/eng/about/bp>
- Statistics Canada. (2016). *Guide to the Survey of Employment, Payrolls and Hours, 2016*. Statistics Canada catalogue number 72-203-G.
- Xing, Y., Lee, B. and Lavallée, S. (2016). *Alternative calculations and extra conditions for aggregation, imputation and outlier detection to improve the SEPH production*. Statistics Canada internal document.
- Yeung, A., Houle, A.-M. and Wirth, S. (2008). "The Challenges of the Use of Administrative Data in the Survey of Employment, Payrolls and Hours." *Proceedings of Statistics Canada Symposium 2008, Data Collection: Achievements and New Directions*. Statistics Canada, Ottawa, Canada.