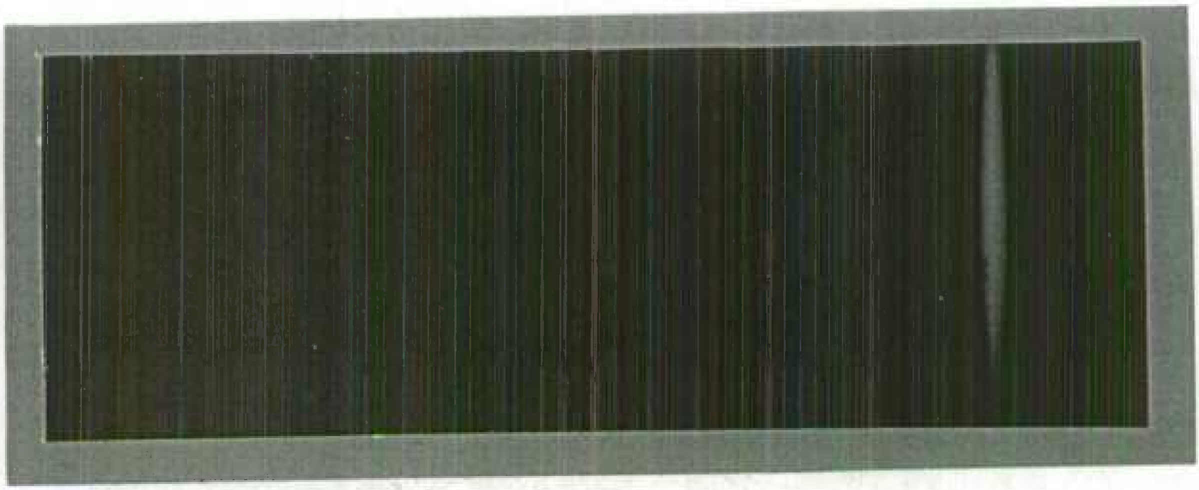
Statistics Canada Statistique Canada



Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes-entreprises

Canadä

WORKING PAPER NO. BSMD-86-007E

CAHIER DE TRAVAIL NO. BSMD-86-007E

METHODOLOGY BRANCH

MÉTHODOLOGIE

ASSESSMENT OF QUALITY OF PAYDAC FILE AND PD20 FORM FRAME DATA

by
Robert Lussier and Desmond Beckstead

June 1985

* This is a preliminary version.
Do not quote without author's permission. Comments are welcome.

Table of Contents

Page

ASSESSMENT OF QUALITY OF PAYDAC FILE AND PD20 FORM FRAME DATA

Robert Lussier and Desmond Beckstead

ABSTRACT

The PAYDAC file and the PD20 Forms are two administrative sources that have been supplied by Revenue Canada-Taxation to Statistics Canada and have been used to update the Business Register (BR).

This paper presents quality issues with respect to the usage of these two sources in the maintenance of the BR. It analyses and summarizes readily available documentation and counts relevant to these issues. Finally it formulates suggestions for future investigations.

The document assumes the reader is generally familiar with the files and system under discussion. These background is provided in the report itself; however, some references are listed which may be used for prior reading if need be.

ÉVALUATION DE LA QUALITÉ DES DONNÉES DE BASE DE SONDAGE CONTENUES SUR LE FICHIER PAYDAC ET LA FORMULE PD20

Robert Lussier et Desmond Beckstead

RÉSUMÉ

Le fichier PAYDAC et les formules PD20 sont deux sources administratives que Statistique Canada reçoit de Revenu Canada-Impôt et utilise pour mettre à jour le RE (registre des entreprises).

Dans ce document, on formule des questions sur la qualité de ces deux sources administratives par rapport à l'usage qu'on en fait dans l'entretien du RE. On y analyse et résume de la documentation et des comptes déjà disponibles sur ces questions. Finalement, on y formule des suggestions pour des recherches futures.

On suppose dans le document que le lecteur est généralement familier avec les fichiers et systèmes discutés. Donc, on ne fournit aucun renseignement de base dans le rapport lui-même; cependant on y liste des références qui pourraient servir de lecture préliminaire si nécessaire.

1. Introduction

Methodology's involvement in assessing the quality of frame data for the Infra-Structure Project was outlined in [1] in three phases. This report presents the result of the first phase, i.e. the initial one, for the PD frame data. This phase consisted in acquiring and summarizing readily available documentation on the evaluation of the quality of the administrative source data.

Two PD data sources are covered in this report. Chapter 2 deals with the PAYDAC file while Chapter 3 covers the PD20 Form. Concise documentation of these sources was recently produced[2]. The authors have used this documentation as the basis of their assessment. They also assume that the reader is generally familiar with the documentation on the files and systems in question.

Each of the two above mentioned chapters describes i) the quality issues; ii) what is known about each issue; and iii) what should be investigated in Phase II and III of the assessment project.

2. PAYDAC File

2.1 Coverage

2.1.1 Issues

Target Population

The authors have defined the target population of the PAYDAC file with respect to the past and current use of the file in Business Register Division (BRD). This decision was made because the purpose of the report is to summarize readily available documentation on the quality of the PAYDAC and the documentation corresponds to its past and current usage. Hence, the authors found it inappropriate at this stage to introduce a definition that corresponds to the potential usage of the PAYDAC file in the Central Frame Data Base (CFDB). This aspect has been left to future phases of the evaluation as explained in the Recommended Future Studies Section of this chapter.

With this introduction in mind, the authors propose that the target population of the PAYDAC file be taken as:

"All businesses which have employees at one point in time during a given calendar reference month in an industry other than agriculture, fishing and trapping, private household services, religious organizations and military services".

The word "business" refers to any corporation or unincorporated analogue of the corporation, including individuals, non-profit organizations and government institutions.

The monthly time reference comes from the fact that the PAYDAC file is the main data source with the Area Frame for birthing new statistical units for monthly surveys.

The above definition of the target population corresponds to the SEPH target population with the exception of the time reference. SEPH conceptual time reference period is the last seven days in the reference month while the above definition refers to a whole calendar month.

"In-scope industries" hereafter denote all industries other than agriculture, fishing and trapping, private household services, religious organizations and military services.

Population Unit Versus Reference Unit

The above definition uses the business as the population unit. However, the PAYDAC reference unit is the Payroll Deduction account (PD). There is one record per account on the file. The PD represents a group of one or more employees for which the business remits source deductions for income tax and employer's and employee's share of contributions to unemployment insurance and Canada Pension Plan. It may or may not correspond to the whole business depending whether the business uses one or more PD's to cover all its employees eligible for these deductions(1). Small businesses generally use only one PD while

(1) This last point raises the question as to whether or not it is possible to have employees that are not eligible to the above mentioned deductions. The authors had no response to this valid question.

large businesses usually have more than one PD. These multi-PD businesses are covered by more than one record on the PAYDAC file. Since there is no cross-referencing of PD's held by the same employer nor any indication that a PD is the sole account of a particular employer, there is a duplication of businesses on the PAYDAC file.

Undercoverage

Suppose now that records for multiple PD businesses could be collapsed into a single record. Each record on the PAYDAC file would then correspond to a complete business. Nevertheless, the file would suffer from both under and overcoverage.

Undercoverage will be caused by individuals or corporations in an in-scope industry who have employees during the reference month but do not have at least one PD account(1). This will happen in four cases. The first one corresponds to new businesses who started operations and hired employees but have not phoned in to RC-T to get a PD account number. The second one corresponds to businesses who surrendered or applied to surrender their PD's but re-started to hire employees without having yet notified RC-T. In this latter case, the PD is still on the PAYDAC file if it is either open or closed for less than twelve months. If the PD is still on the file, then it is simply a misclassification problem (it would be coded as administratively maintained or closed when it is in fact open). Otherwise, it is really undercoverage. The

(1) A business is said to have a PD account when it has a PD account on the PAYDAC file. The account may be open, closed or administratively maintained as explained in the PAYDAC/PD Processing Documentation[2].

third case corresponds to businesses who are coded to an out-of-scope SIC when in fact they belong to an in-scope industry. A fourth case exists when a PD account is shared by more than one company; the situations within which this could happen are:

i) within an enterprise (i.e. the companies are associated through ownership);

ii) within a joint venture (i.e. the companies are associated through a contract);

iii) as the result of a business sale and purchase where the PD is taken over by the new owner.

When a business asks for a PD account, the account number they obtain is added to the PAYDAC File in a month which is not necessarily the month during which they first hired employees. RC-T cannot do any better as the PAYDAC file is continuously updated. It does not have a reference period other than the calendar year, which is the fiscal year for individuals. The monthly reference period is a request by STC as the PAYDAC is used as the main source for birthing businesses for the monthly surveys.

Overcoverage

Overcoverage can be caused by three different situations. The first one corresponds to individuals or corporations who have no employees during the reference month but have a PD account. It is impossible to know directly from the PAYDAC file whether the business had employees or not during the month. The business may not have employees for two reasons. It may be a new business that have not started to hire employees although they obtained a PD or they may be an on-going business who stopped employing people for at least the reference month. One may wonder in these cases whether the business is really out-of-scope. One may argue that it should remain in-scope if it is not known whether or not the business will be an employer in the future. The argument could be based on the difficulties of birthing and deathing the same business from one month to the next depending on its employer status and on the length of time that a central frame takes to react through administrative files to a change in the real world. Nevertheless, everybody would agree that a business is definitely out-of-scope for PAYDAC when it stopped operations and have no plans to resume them. However, the PD account will remain open if the business does not apply to surrender the PD. The only way of closing an account, other than the business requesting it, is through ad hoc enquiries initiated by RC-T. Periodically, RC-T undertakes contact with holders of PD accounts which have not had any account activity. Accounts may be closed as a result of these contacts.

The second situation causing overcoverage corresponds to individuals or corporations who have employees during the reference month, have a PD account but belong to an excluded industry. However, these may not represent real overcoverage. If they are correctly SIC coded, they can be excluded from tabulations and frames. They would nevertheless represent wasted efforts as long as their industry is excluded from the coverage of the BR.

The last situation leading to overcoverage concerns businesses who are coded to an in-scope SIC when in fact they belong to an excluded industry.

2.1.2 What is Known about the Issues

Characteristics of Records on PAYDAC File

The number of records on the PAYDAC file is presented in Table 1 for a recent 12 month period by status of account and by current year balance being equal to zero or not. The corresponding percentages are presented in Table 2.

Table 1

Number of Records on the PAYDAC File for each BRMF Cycle
by Status of Account and by Current Year Balance
Being Equal to Zero or Not

BRMF Cycle Number	Date PAYDAC Processed	Open Account		Closed Account		Total
		C.Y.B.≠0	C.Y.B.=0	C.Y.B.≠0	C.Y.B.=0	
151	2/5/84	588,912	399,166	0	324,691	1,312,769
152	7/6/84	618,413	381,296	0	332,750	1,332,459
153	5/7/84	651,910	360,386	0	338,859	1,351,155
154	2/8/84	686,565	334,675	0	347,475	1,368,715
155	10/9/84	708,652	319,170	0	354,439	1,382,261
156	3/10/84	721,766	305,682	0	365,978	1,393,426
157	5/11/84	736,213	287,136	0	139,673	1,163,022
158	5/12/84	746,797	277,349	0	150,733	1,174,879
159	7/1/85	751,772	276,387	0	154,790	1,182,949
160	14/2/85	24,952	1,015,076	0	159,700	1,199,728
161(1)	4/4/85	586,588	467,401	0	179,663	1,233,652

(1) Cycle 161 could not be run in March 1985 because RC-T was not able to provide a clean input tape. Therefore cycle 162 covers 2 months of updates.

Table 2

Percentage of Records on the PAYDAC File for each BRMF Cycle by Status of Account and by Current Year Balance Being Equal to Zero or Not

BRMF Cycle Number	Date PAYDAC Processed	Open Account		Closed Account		Total
		C.Y.B.≠0	C.Y.B.=0	C.Y.B.≠0	C.Y.B.=0	
151	2/5/84	44.86	30.41	0	24.73	100
152	7/6/84	46.41	28.62	0	24.97	100
153	5/7/84	48.25	26.67	0	25.08	100
154	2/8/84	50.16	24.45	0	25.39	100
155	10/9/84	51.27	23.09	0	25.64	100
156	3/10/84	51.80	21.94	0	26.26	100
157	5/11/84	63.30	24.69	0	12.01	100
158	5/12/84	63.56	23.61	0	12.83	100
159	7/1/85	63.55	23.36	0	13.09	100
160	14/2/85	2.08	84.61	0	13.31	100
162(1)	4/4/85	47.55	37.89	0	14.56	100

(1) Please refer to footnote (1) of Table 1.

It can be seen that:

a) The number of closed accounts is fairly large. It represents between 12 and 27% of the total number of records on the file. It increases from month to month until cycle 157 for which RC-T purged some closed accounts. There were over 200,000 PD accounts that were removed from the PAYDAC file because they had been closed for at least 12 months and no reactivation had occurred.

b) The number of remittances field in the first month of the year is very low. This supports the statement that there is a lag between the reference month of the remittances and the month during which the remittances are posted on the PAYDAC file.

Population Unit Versus Reference Unit

There are no tables currently available from the PAYDAC file which use the business as the tabulation unit; all use the PD number instead. So there is no precise way of showing the duplication of businesses on the PAYDAC file which results from the multi-PD businesses being reported for each of their PD's. However, the number of mode 3 authority 03 records on the Business Register Master File (BRMF) is used below as an indicator of the situation pending future investigations.

The mode 3 authority 03 are PD-defined records. These records are created when a business has at one point in time at least 2 PD's or when the PAYDAC file contains frame data different from a survey for a single-PD business. Each second and each subsequent PD of a multi-PD business on the PAYDAC file corresponds to one and only one mode 3 authority 03 record on the BRMF if the BR linkage operations worked correctly. Let us assume that they did. Also, the first PD may or may not correspond to a mode 3 authority 03 depending on the situation. On the other hand, each mode 3 authority 03 record on the BRMF corresponds to one and only one record on the PAYDAC file. This record may be for a single or a multi-PD business. Thus the number of mode 3 authority 03 cannot be used as the number of PD's belonging to a multi-PD business.

The number of mode 3 authority 03 records is presented in Table 3 by cycle. It gives an indication of the complexity of the current BRD operations in response to using the PAYDAC file with the business at the population unit.

Table 3

Number of Mode 3 Authority 03 Records on BRMF by Cycle

BRMF Cycle Number	Number of Mode 3 Authority 03			Total
	Active	Administratively Maintained	Closed	
151	217,421	33.659	107,058	358,138
152	221,897	31,730	110,788	364,415
153	224,601	31,644	113,684	369,929
154	228,296	31,215	116,938	376,449
155	229,611	31,925	119,603	381,139
156	234,300	30,256	123,808	388,364
157	237,243	29,712	49,948	316,903
158	241,342	29,118	54,348	324,808
159	245,343	29,110	56,406	330,859
160	237,353	42,159	59,346	338,858
162[1]	227,185	57,200	67,964	352,349

[1] Please refer to footnote (1) of Table 1.

Identification of the Births

Currently, each time a new PD is found on the monthly PAYDAC file, it is birthed as an unclassified record on the BRMF. It then takes some time before it is found whether or not the PD corresponds to a new potential employer business and then whether the new potential employer business operates in an in-scope industry.

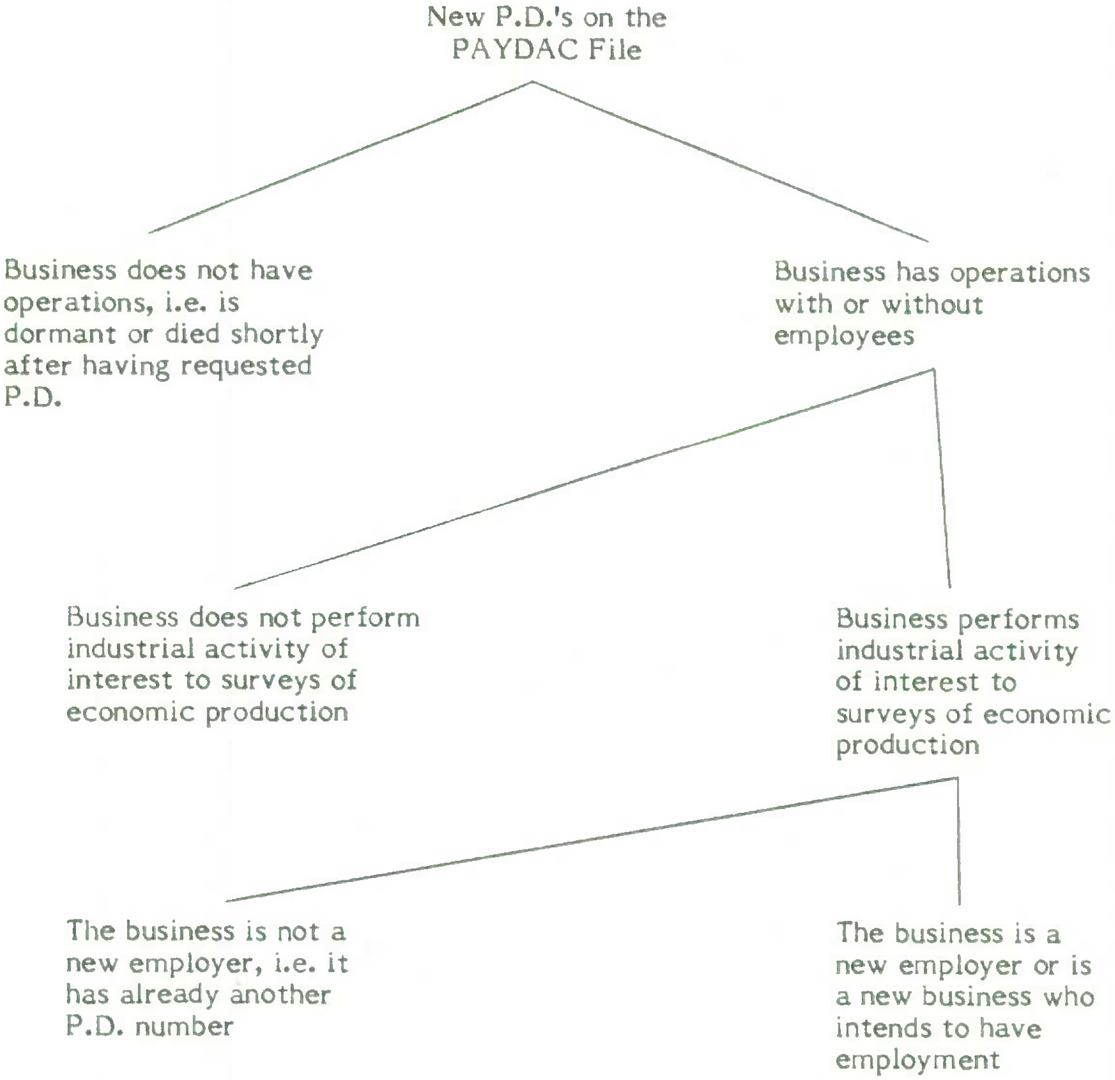
These two problems correspond to the population unit versus PAYDAC reference unit problem and to the overcoverage caused by excluded industries problem but applied specifically to births. There exists a study which measured in a certain way these problems. It is referred to as the PD Processing Time Lag Study[3]. However, its results do not distinguish the two problems as they are explained immediately above and in the issues section. The framework that was developed for the study is different. It looks at whether or not the business is operational in an in-scope industry before checking if it is a new potential employer business. This framework is presented in Figure 1. It illustrates a classification scheme for PD accounts.

The Study estimated the proportion of new PD's on the PAYDAC file acquired in March 1984 which fall into each of the classes of the framework. These proportions are reported in Figure 2. The Study could not however resolve the case of 17.8% of the new PD's that is to say that it could not decide whether these PD's corresponded to new potential employer businesses or to businesses who had already another PD.

The reader should also note that there was a lag between the time the births were identified from the PAYDAC file and the time the survey data were collected. It is thus possible that some births were new employer businesses who operated in an in-scope industry before the data were collected at which point they had stopped their operations. Similarly, it is also possible that the new PD belonged to a business who was bought before being enumerated for the Study by another business who is a new employer. The buying business may use the PD of the business they bought. If this was the case then in the Study, the PD was classified as a non-operational legal entity because the selling business was the only one contacted even though the PD may now correspond to a new employer business.

Figure 1

Conceptual Framework for the Classification of New P.D.'s



New P.D.'s on the PAYDAC File

Business does not have operations, i.e. is dormant or died shortly after having requested P.D.

Business has operations with or without employees

Business does not perform industrial activity of interest to surveys of economic production

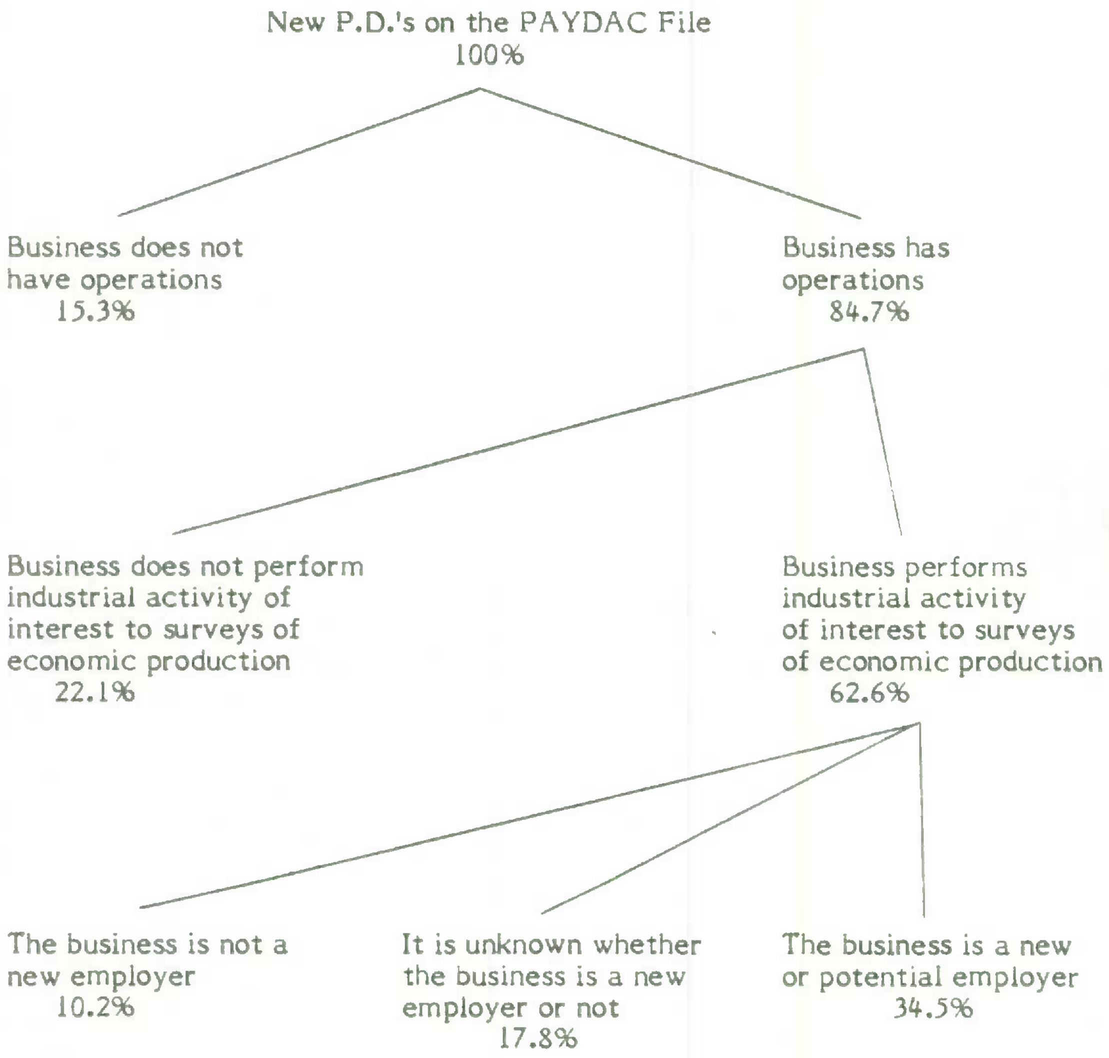
Business performs industrial activity of interest to surveys of economic production

The business is not a new employer, i.e. it has already another P.D. number

The business is a new employer or is a new business who intends to have employment

Figure 2

Estimates of the Proportion of New P.D.'s Falling in Each Class of the Conceptual Framework for February 1984 Reference Month


New P.D.'s on the PAYDAC File 100%
Business does not have operations 15.3%
Business has operations 84.7%
Business does not perform industrial activity of interest to surveys of economic production 22.1%
Business performs industrial activity of interest to surveys of economic production 62.6%
The business is not a new employer 10.2%
It is unknown whether the business is a new employer or not 17.8%
The business is a new or potential employer 34.5%

Notwithstanding the above, the percentages presented in Figure 2 remain very interesting. They show that about half of the new PD's end up as not being new potential employer businesses who operate in an in-scope industry. The reader does not need to assume that there is no variation between months to extrapolate the data. The magnitude of the problem is striking.

There is another limitation to the study data. It refers to actual and potential employers while the target population refers to actual employers. Thus the real percentage of PD's who correspond to actual new employer businesses is most likely less than 50%.

Undercoverage Caused by New Businesses who Hired Employees Before Getting a PD Account

The PD Processing Time Lag Study provides measures for this undercoverage. The Study consisted in surveying for various date characteristics a sample of PD births added to the BRMF in March 1984 from the PAYDAC file. It is different from the date the business obtained its PD by two lags. The first one is the lag between the time that the business obtained its PD to the time RC-T created the monthly PAYDAC for STC. The second one is the lag between the creation of the monthly PAYDAC file to the time BRD runs it against its BRMF. Tables 4 and 5 were built for this report from the PD Processing Time Lag Study results with the assumption that these two lags take together cover one month.

Table 4 estimates that 3-4 months after having been operational with or without employees, 50% of the new businesses who requested/will request a PD are on the PAYDAC file. This is different from saying that 50% of all new businesses are on the PAYDAC file within 3-4 months after they started operations. The difference resides in the businesses who will not get onto the PAYDAC because they will not request a PD number. The PD Processing Time Lag Study could not cover those and therefore they are excluded in the estimates.

Similarly Table 5 estimates that 1 month after having hired employees, 50% of the new businesses who requested/will request a PD are on the PAYDAC file. A result from Table 5 is however surprising. One would have thought that there is an implicit upper limit for the time a business can have employees without being registered with RC-T. The registration for remittances of income tax, unemployment insurance premiums, etc., is mandatory by law. It is thus surprising to see that it took 5 to 6 months before 75% of the new businesses with employees got onto the PAYDAC.

Overcoverage Caused by Businesses with a PD but No Employees

There are some data on this subject from the PD Processing Time Lag Study and from printouts produced by the BRD operations.

First, the PD Processing Time Lag Study estimates that 15.3% of the business that get a PD, whether their first one or an additional one, were not operational at time of survey, i.e. 3 to 5 months after their PD account was added to the PAYDAC file. Also, the Study estimates that 23% of the new businesses that got a PD, were operational and were in

Table 4

Domain: New Businesses who Perform an Industrial Activity of Interest to Surveys of Economic Production and will Eventually Get on the PAYDAC File

Estimate of the Number of Months After a Set of Domain Units is Operational to Get Various Percentages of this Set on the PAYDAC File

Estimate of the Number of Months	Percentage on the PAYDAC File
1 month	25%
3-4 months	50%
17 - 21.5 months	75%

Table 5

Domain: New Employer Businesses who Perform an Industrial Activity of Interest to Surveys of Economic Production and will Eventually Get on the PAYDAC File

Estimate of the Number of Months After a Set of Domain Units Hired Employees to Get Various Percentages of this Set on the PAYDAC File

Estimate of the Number of Months	Percentage on the PAYDAC File
0 month	25%
1 month	50%
5 - 6 months	75%

an in-scope industry had not hired yet employees at time of survey. These figures are again based on the assumption that it takes one month between the time a business request its PD from RC-T and the time the PD is added onto the BRMF.

Secondly, counts from the BRD operations show that there is a certain number of PD births that have already a closed status or an administratively maintained status when they first appear on the monthly PAYDAC file. The exact counts are presented in Table 6. These closed and administratively maintained births are nevertheless currently birthed as unclassified records on the BRMF.

Thirdly, Tables 1 and 2 of this report show that the number of open accounts with current year balance (CYB) equal to zero is fairly large. Excluding cycle 160, it represents between 21 and 38% of the records on the monthly PAYDAC. This indicates that there are many businesses that acquire a PD but do not use it in the same calendar year. These figures may however be overestimates (most likely small) of the situation. A business that has been a non-employer since the beginning of the calendar year has a CYB equal to zero but a CYB equal to zero does not mean the business has been a non-employer since the beginning of the calendar year. It could be a late remitter. Late remittances after the calendar year end are not on the PAYDAC file that BRD gets. That is to say that the CYB is set back to zero on the PAYDAC file that we get when a new calendar year starts. If a business remits after the new year for the previous year, then the amount is posted for the previous year. BRD does not get a value for these remittances. Thus,

Table 6

Table of PAYDAC Births by Registration Status Code

Cycle	Active	Administratively Maintained	Closed	Total
151	15,442	57	62	15,561
152	19,532	58	100	19,690
153	18,562	27	107	18,696
154	17,449	20	91	17,560
155	13,417	28	101	13,546
156	11,089	21	55	11,165
157	13,063	25	99	13,187
158	11,751	14	92	11,857
159	8,041	9	20	8,070
160	16,098	607(2)	74	16,779
162(1)	15,079	18,237	608(2)	33,924

(1) Please refer to footnote (1) of Table 1.

(2) BRD could not get any explanations from RC-T of the reasons for these very high figures.

among the 276,387 PD's which were classified as open with a CYB equal to zero in cycle 159, there is probably a small proportion that submitted late remittances in the following months for what was then the previous year.

2.1.3 Recommended Future Work

Target Population

The authors have given at the beginning of Section 2.1.1 a definition of the target population of the PAYDAC file with respect to its past and current usage in BRD. However, a new target population for the PAYDAC file was implicitly given in the Information Model and the Data Model developed recently for Prototype 1 of the CFDB. In these models, the PD's are attached to elements of the operating structure and not to legal entities as is currently done. The findings reported in this paper definitely supports this initiative. Thus the following is recommended for the second phase of the evaluation.

1. A definition should be written for the elements of the operating structure to which the PD's would be attached.

2. A decision as to whether the PAYDAC is used to identify elements of the operating structure for employers, potential employers or any business should be made. This is an important decision as it would give the direction to the next phase of the assessment as well as to the development of the specifications for the processing of the PD inputs for the various prototypes.

3. The target population of the PAYDAC should be revised in light of the above decision and in light of the elements of the operating structure to which the PD's would be linked.

Population Unit Versus Reference Unit

A Study referred to as the PD Linkage Study is being initiated by sub-team 3M of the Infra-Structure Project. It will provide various data that will be relevant to Phase 2 of the evaluation of the PAYDAC file. Among others, it will calculate the number of businesses by number of associated PD accounts. This tabulation will be derived from the BRMF and will define businesses as active reference units on the BRMF.

Undercoverage

The authors propose that the following be done with respect to undercoverage.

1. A proposal should be written on how to estimate the impact in terms of economic activity of businesses that do not get on the PAYDAC file because they do not request a PD. These would be covered by tax for the annual surveys but would not be covered for the sub-annual surveys if they are not in the integrated portion of the CFDB.

2. A proposal should be developed on how to calculate the time lag between the time a business applies for a PD and the time it is added onto the BRMF. This report made the assumption that there is a time lag of one month but this should be verified.

3. An investigation should be made on how more frequent can we get the PAYDAC file from RC-T together with the impact on cost, resources, etc., for each additional number of times a PAYDAC file is received.

Overcoverage

The authors propose that the following be done:

1. The question of whether PD should be birthed or not when they have already a closed or administratively maintained status should be studied in more detail. The possibility of birthing PD's only when they get the active status and are not already on the BRMF should be considered. This means that PD's would be added onto the BRMF.

   a) when they are birthed as active (open and not administratively maintained) on the PAYDAC file; or

   b) when their status on the PAYDAC file changes from inactive or administratively maintained to active and they had not been previously added onto the BRMF. The study should also consider the time lags involved in these changes.

2. The following tables which give more information on births should be produced.

a) the proportion of active, administratively maintained and closed accounts for various lengths of time after the accounts were birthed onto the PAYDAC file.

b) the proportion of accounts with CYB = 0 and CYB ≠ 0 for various lengths of time after the accounts were birthed onto the PAYDAC file, including how long it is to the first remittance.

3. If the target population refers to operating structure elements that have employees (and not that are potential employers), then an investigation should be made on using the CYB being equal to zero or not as a mean for birthing and deathing employers.

2.2 Response Rate

2.2.1 Issues

No Data in Mandatory Fields

There are fields on the PAYDAC file for which a non-blank content is always expected. These are generally fields that identify the employer and determine its status. These fields are the following:

1. PD Account Number
2. Name of employer (line 1)
3. Name (line 2)

4. Address line 1
5. Address line 2
6. Address line 3
7. Address line 4
8. Employer's postal code
9. Language code for output
10. PAYDAC Master type
11. Activity code .

Although a non-blank content is always expected for each of these fields, it is theoretically possible that some of them be occasionally left blank. (The exception to this rule may be some name and/or address lines.) These blank fields would be a source of concern in the processing of the data by Statistics Canada.

Difficulty in Identifying Non-Respondents to the CYB Field

The CYB field is expected to be non-blank when the business had employees eligible for deductions in the current calendar year. The amount in absolute value in the CYB field is expected to grow from month to month as the business files remittances. However, there is no way of knowing whether an increase in a given month corresponds to a remittance for this reference month or to a late remittance (i.e. a remittance for a previous reference month). Similarly, no increase does not necessarily mean a non-response to the reference month: the business may not have had employees during that month. Thus there is no easy way of identifying a non-respondent to a given reference month and no easy way of calculating response rates. All decreases in CYB (except for year end) are significant.

2.2.2 What is Known About the Issues

No Data in Mandatory Fields

No tables of response rates for the mandatory fields were readily available. They have to be produced.

Difficulty in Identifying Non-Respondents to the CYB Field

The report of the PD7 Analysis Team of the Employment Statistics Development Project [4] alludes to remittances carrying a reference month in the PAYDAC system at RC-T. The RC-T PAYDAC system contains more data than the PAYDAC file obtained so far by Statistics Canada and documented in [2] by this project. The STC PAYDAC file is a file created from the PAYDAC system by RC-T specially for STC. Nevertheless, the report said that as of 1979, the RC-T PAYDAC system placed a higher priority on getting the deduction monies into RC-T than on keeping track of to exactly which month a given payment referred. If this statement still applies, it does not appear very possible to calculate response rate for the CYB for a survey-taken reference month even if one had access to the RC-T PAYDAC system.

A more recent study, the PD-Processing Time Lag Study, provides response pattern information about the CYB for births. Table 7 which is produced from the Study shows the percentage of records in various CYB ranges for various months after birth on the PAYDAC file. During the first month, i.e. in February 1984, 78.6% of the February births had a zero CYB. Three months after, i.e. in May 1984 only 33.6% of the February births still had a zero balance. However again, it is not known

Table 7

Percentage of Births in Various CYB Ranges by Number of Months After Birth on PAYDAC File

Months After Birth on PAYDAC File	CYB					
	Zero	-0.01 to -200.00	-200.01 to -500.00	-500.01 to -1,000.00	≤ 1,000,000	Total
0 (Feb. 84)	78.60	10.60	4.53	2.67	3.60	100.00
1 (Mar. 84)	44.87	23.13	13.53	8.27	10.20	100.00
2 (April 84)	37.00	19.00	14.00	11.67	18.33	100.00
3 (May 84)	33.60	15.67	13.47	12.26	25.07	100.00

to which reference month these remittances apply as well as whether the zero CYB businesses are non-respondents or simply non-employers.

2.2.3 Recommended Future Work

No Data in Mandatory Fields

The authors recommend that response rates be calculated for the non-blank mandatory fields listed in Section 2.2.1 above.

Difficulty in Identifying Non-Respondents to the CYB Field

The documentation [2] presents the CYB as a potential classification field for the CFDB. It then becomes important to make efforts to calculate response rates for this field before it is used for this purpose. The points reported in the previous section from the PD7 Analysis Team Report should then be checked. The analysis is several years old. It is known that the RC-T PAYDAC system has changed since then. More specifically, the following questions should be answered.

1. Whether or not remittances carry a reference month at RC-T?

2. If yes to 1., whether or not the reference month is reliable, i.e. whether or not RC-T really keeps tract of to which month a given payment refers?

If the answers to these questions are positive, then a proposal should be developed on how response rates would be calculated. The proposal should show the cost and resources involved in obtaining data by reference month from RC-T.

In addition to being of use to calculate response rates, the remittances by month may have some potential as a classification variable or as an imputation model variable. In the later case, surveys would use the monthly remittances when imputing for partial or total non-response. Thus, if answers to questions 1 and 2 above are positive, a proposal should be written showing recommended usage of the monthly remittances along with the cost and resources involved to acquire them monthly and to put them on the CFDB.

2.3 Other Data Quality Problems

2.3.1 Issues

RC-T PAYDAC System Documentation

What has been documented so far by this project is the PAYDAC file obtained by STC [2]. This is different from the PAYDAC system maintained by STC as explained in Section 2.2.2 above. The PAYDAC system may contain data that STC does not currently obtain but may wish to acquire in the future. This is an appropriate timing for addition/deletion of fields as the central frame system is being redesigned. However, there is no up-to-date integrated documentation of the PAYDAC system available in STC.

Use of Current Year Balance to Estimate Employment

As explained in the previous section, remittances paid into (and T5 deductions from) a PD account are indicated by the Current Year

Balance (CYB) field on the PAYDAC file. In spite of the occurrence of irregular remittances and T4's which have been filed early, it was thought that CYB data could possibly be used regularly to estimate employment size codes for PD-defined units.

Province of Employment Fields

The PAYDAC system allows for multi-provincial use of a PD account; in other words, an employer may remit to a single PD account monies collected from employment occurring in more than one province/territory. The issues here are twofold:

i) the extent of the use of this flexibility, and

ii) whether or not most of the employment covered by the account is in the same province as the employer.

Thirteen fields are available for the indication of employment: one for each province or territory and one for foreign.

Number of Employees Field

The number of employees field is believed to be defined as the number of T4 Supplementaries indicated on the T4 Summary of the previous year. This field has the following problems:

1. Births do not have previous year T4 Summary and thus do not have a non-zero number of employees.

2. Some employees receive more than one T4 Supplementary which leads to some overcounting of employees.

3. Up to and including 1978, this field was set to zero at the beginning of each year. As a result, PD accounts which submitted late T4 Summaries (i.e. later than December in the current year) have a zero in the Number of Employees field until a T4 Summary is received. The current procedure is not known.

Activity Code Field

If a business with an active PD account informs RC-T that it will no longer be carrying out an employer activity under this account then, within the PAYDAC system, the activity code is switched from 0 to 1. This value of 1 could be interpreted to mean that the account is open for administrative use only. The issue concerns the reliability of this field as an indicator of non-employer activity.

Response Error in PAYDAC Name and Address Fields

Concerning 'name', the issue is that PAYDAC may have either the legal name or the operating name. This name, in either case, may be correctly spelled, have misspelling(s), or be out-of-date (i.e. a change of name has occurred).

Concerning 'address', the issue is that the PAYDAC address may refer to either the physical location of the economic activity conducted by the business or the mailing address or neither location nor mailing address.

Concerning the postal code, the issue is that it may or may not correspond to the address, given whether the address is for the location or for mailing.

2.3.2 What is Known About the Issues

RC-T PAYDAC System Documentation

RC-T has converted the PAYDAC system to a data base environment since the PD7 Analysis was conducted. More recently, changes seem to have been made to the system as RC-T could not provide cycle 161 to STC. Characteristics of the births identified in cycle 160 and 162 (refer to Table 6) appear to be quite different from previous cycles.

Use of Current Year Balance to Estimate Employment

During fiscal year 1983-84 Labour Division sponsored a size coding test project involving itself, BSMD and BRD. A test system was developed to estimate employment size codes based on a model which used as input:

a) CYB data from the monthly PAYDAC files,

b) annual wage and salary and remittance data from the T4 Summary file, and

c) average weekly earnings data from the monthly Survey of Employment, Payrolls and Hours (SEPH).

These estimated employment size codes were then compared within the test system to size codes based on data reported on SEPH survey questionnaires. It was found that by adjusting the estimated employment to compensate for irregular remittances to the PAYDAC system, that the estimated SEPH size codes agreed with the codes based on reported employment for about 82% of the employment reporting units tested. This model underestimated size codes for approximately 6% of the units tested.

In adjusting the 'raw' estimate to compensate for irregular remittances, several techniques were tested. One of these techniques used data from the T4 Summary File for Number of T4 Supplementaries. (This data is obtained by the PAYDAC system in order to update the Number of Employees field annually.) The technique using the Number of T4 Supplementaries maximized the number of test units which had their estimated SEPH size code equal to that which was based on their reported SEPH employment data.

Province of Employment Fields

The PD accounts which were birthed within the PAYDAC system during the month of February 1984 were monitored for use of these fields. It was found that within the first quarter of a year after their birth, less than 5% of these new PD's indicated employment in more than one province.

Number of Employees Field

The PD7 Analysis Team Report says that many accounts which appear to have zero employees do indeed have employees, Businesses could have 0 in the Number of Employees field but have a non-zero value in the CYB. The Analysis, although outdated, reported that there were about 100,000 such accounts in the 1977 PAYDAC file and in the 1978 PAYDAC file.

Since the PD7 Analysis project RC-T has added a question on the PD7AR asking for the number of employees. However, the responses to this question are not passed on to STC.

Activity Code Field

A study was conducted by BRD in 1981 to assess the quality of this field. It was found that 93-94% of the sampled units on the PAYDAC file having an open account and an activity code of 1 did in fact have no employer activity. The remaining 6-7% of the sampled units were cases of:

i) not satisfying the STC/RC-T definition of employment, or

ii) business reorganization, or

iii) incorrect setting of the code.

All closed accounts have an activity code of 1.

Until recently, all accounts birthed within the PAYDAC system have had a code value of 0. However, beginning with BRMF cycle 162, the majority of PD births have appeared with an activity code of 1.

Response Error in PAYDAC Name and Address Fields

A joint study (called SARUS) involving BRD, BSMD and the STC Regional Offices was recently conducted on a sample of single active reference units on the BRMF that were classified as being in-scope for surveys of economic production.[5] Several estimates from the study are presented here for the domain which is the set of units under the same operational legal entity at one location for the same establishment. When the name obtained in the interview was compared with the name on the BRMF, three outcomes were considered by the study:

i) name requires no correction

ii) name has a misspelling

iii) name has changed (other than misspelling).

The proportions of the domain estimated to belong in each of these categories were as follows:

i) no correction required: 93.33%

ii) misspelling: 1.16%

iii) change (other than misspelling): 5.51% (with a CV of 7.68%).

When the CBS-718 address collected in the interview was compared with the BRMF address (usually obtained from the PAYDAC file), three outcomes were considered by the SARUS study:

i) the PAYDAC/BRMF address was for the location of the establishment (and may also have been the mailing address).

ii) the PAYDAC/BRMF address was only for mailing.

iii) the PAYDAC/BRMF address represented neither the location of the establishment nor a valid mailing address for contact.

The relative occurrence of these outcomes were estimated to be as follows:

i) address is location: 67.31%

ii) address is only for mailing: 23.91%

iii) address is neither location nor mailing: 8.78% (with a CV of 6.05%).

When the postal code for the establishment's location (collected in the interview via the CBS-718) was matched to the STC postal code conversion file, the resulting Standard Geographical Codes (SGC's) were compared to those on the BRMF; these comparisons were done at the 2-digit (province) and 7-digit levels (Census sub-division/municipality). Proportions of the number of units in the domain were estimated to be:

a) change in province: 1.40% (with a CV of 14.76%);

b) change in 7-digit SGC: 8.59% (with a CV of 5.51%).

2.3.3 Recommended Future Work

RC-T PAYDAC System Documentation

An up-to-date documentation of the RC-T PAYDAC system should be obtained and always maintained. As a start in the short term, the Operational System Section and the Computer System Section of the PD7 Analysis Team Report [4] should be updated.

Use of Current Year Balance to Estimate Employment

A production Size Coding System is currently being developed for the joint use of BRD and Labour Division. It is recommended that the documentation that has been prepared on this system be reviewed for the deliverables that will be produced. Any requirements of Sub-project Team 1 that will not be met should then be identified.

Province of Employment Fields

For the purposes of geographical statistics, STC creates special 'provincial' establishments by subdividing regular establishments. Multi-provincial establishments may continue to exist, however, if the data required to allocate the value added is not made available. Hence it is recommended that all active PD accounts which are currently in use be investigated for multi-province employment. Stratification should be done by age of account, with emphasis on multi-region participation.

Number of Employees Field

The authors recommend that the following be done.

1. A cross-tabulation of the "Number of employees" field with the "CYB" field should be produced for some recent monthly PAYDAC files to check if the findings of the PD7 Analysis Team still apply.

2. RC-T should be contacted to verify the procedures used for determining the number of employees (e.g. whether or not they still use the T4 Summary; whether or not they set the field to zero at the beginning or end of each year).

3. A feasibility study should be done on using the number of employees from the PD7AR rather than the number of T4 Supplementaries from the T4 Summary. This feasibility study should show the advantages and disadvantages of each approach as well as the cost and resources involved in implementing the new alternative.

Activity Code Field

The rationale for birthing PD's with an activity code of 1 should be discussed with RC-T; their methodology for generating activity code updates should also be included in the discussions. The impact of recent changes to RC-T procedures related to the activity code field should be analyzed, at least as far as they affect BRD and SEPH operations.

N.B.: This future work has also been covered, in part, in the section on overcoverage (see 2.1.3).

Response Error in PAYDAC Name and Address Fields

The single-establishment active reference units studied in SARUS exist on the BRMF as either PD-defined or bureau-defined records. Name and address data for PD-defined records are only updated by changes observed for these fields on the PAYDAC file; BRD staff update these fields for bureau-defined records. Hence the quality of these fields on the BRMF as measured in the SARUS study does not relate solely to PAYDAC data; (some 25-35% of the units studied were probably bureau-defined).

It is recommended that the data files generated from the study be analyzed further for that subset of the units which are PD-defined. Re-use of the quality measurement programs should be considered in producing the estimates for these units.

3. PD20 - Employer Registration Form

3.1 Coverage

In order to comply with legal requirements, potential employers request Payroll Deduction (PD) account(s) from Revenue Canada - Taxation. RC-T then issues to each such employer a PD account number and the following set of documents.

a) tables of deductions for Unemployment Insurance, Income Tax and Canada Pension Plan,

b) a PD20 (Employer Registration form),

c) a PD7AR (Tax Deduction, Canada Pension Plan, Unemployment Insurance Remittance Return),

d) T4's (Statement of Remuneration Paid),

e) T4 - T4A Summary (Summary of Remuneration Paid),

f) TD1 (Tax Exemption Return), and

g) associated information guides and booklets.

Each new employer is requested (though not legally required) to complete the PD20. It is assumed that the coverage of the PD20 is the same as that for the PAYDAC system since a PD20 form is issued every time that a PD account is opened or birthed within the PAYDAC system. It is not known what RC-T does with a PD20 for which there is no corresponding PD account on the PAYDAC file.

3.2 Response Rates

3.2.1 Issues

Having completed the PD20 (in quadruplicate), new potential employers retain the fourth copy, and mail the first three copies to the nearest RC-T District Office. There the PD20's are separated and the copies destined for Statistics Canada (STC) are batched. The District Offices then mail, on a daily basis, these batches of PD20's to STC. Our internal mail service then delivers these batches to Business Register Division (BRD) where they are unpacked and the PD20's are re-batched in preparation for data capture. After another trip through the STC internal mail service, the forms are key-edited with 100% verification for some data fields. Once the data capture has been completed, the forms are returned to BRD and a notification is sent indicating which specific data file has been created. In parallel with this data capture process, BRD backs up these PD20 tape files on a daily basis. Then on a weekly basis, these data files are processed by the PD20 system; the output of this system takes two forms:

i) documents for clerical follow-up

ii) automated transactions to be executed as of the next Business Register Master File (BRMF) update.

The above description shows that it is normal to expect a time lag between the birth of a PD account on PAYDAC and having a keyed PD20 available for processing in BRD.

If, in attempting to update records on the BRMF, it is found that the PD-defined record has not yet been created (i.e. the PD number has not yet appeared on the PAYDAC file), then the PD20 data record is written out to the Pending file; this file is then used as input to the next BRMF update run. The number of records on this file varies from 200 to 1,000.

Currently, the classification data on PD20 forms are by far the major source of information by which PD-defined records are classified on the BRMF. As a part of the regular monthly monitoring of each BRMF production cycle, counts are available for the number of unclassified (SIC = 70099) PD-defined records by 'date of birth'. Hence the classification process can be monitored from at least the following perspectives:

i) the rate at which a particular month's set of births are classified from one cycle to the next.

ii) given the set of unclassified records for a particular BRMF cycle, what is the age distribution of these records (based on the BRMF cycle in which they were birthed)?

After each BRMF has been created, the file is scanned for active records (i.e. survey status code = 8) which are still unclassified 12 months after they have been birthed (i.e. after having been on the BRMF for 13 months). These records are then matched to the entire BRMF by name with the objective of retaining only those records which correspond to new PD accounts for new businesses. The resulting file of unclassified (unmatched) records is then input to the company contact control system and labels are printed for CBS-718 forms. These 718's are then sent to the STC RO's by Priority Post. The RO staff complete the 718's by conducting telephone interviews with the employers. Having been batched in the RO's, the 718 forms are delivered by Priority Post to Head Office, and again by the internal mail service to BRD. The forms are then rebatched in BRD, sent to data capture and key-edited (with 100% verification of certain fields). The forms are returned to BRD and notification of the resulting data file is issued. Based on the work assignments which have been pre-specified for the next clerical processing month in BRD, a set of these 718 data files are selected to be input to the CBS-718 processing system which is executed once each month. Thus it is possible that some completed CBS-718 forms will not be processed during the month of their receipt. This system edits the 718 data and classifies each 718 according to the type of updates it may generate. The output includes both automated BRMF update transcriptions and printed output for clerical scrutiny.

It should be mentioned that this follow-up activity is not limited to 12-month-old unclassified records. Sometimes a PD20 form is received from an employer, but the data given is insufficient to assign an SIC. At this time another similar type of follow-up is initiated, whereby a card is mailed to the business. This card is used to obtain the information necessary to determine the nature of the business.

In either case, the objective of the follow-up activity is to obtain sufficient data to assign an SIC to the PD-defined record.

Again by using the counts described above for unclassified records, the follow-up scheme can be evaluated.

3.2.2 What is Known About the Issues

Based on the production monitoring counts of unclassified records described above, the following data was obtained:

counts for the number of active 70099 PD-defined records by set of BRMF cycles and quarter after birth, where

'active' means survey status code (SSC) not equal to 8 (inactive).

'70099' refers to the SIC value for an unclassified record,

'PD-defined records' means those records associated with PD accounts and updated based on RC-T data,

'set of BRMF cycles' refers to the three months of BRMFs whose reference months fall within the same calendar quarter; (for example, cycles 154 to 156 have reference months of July to September, 1984), and

'quarter after birth' means the month corresponding to 3 months after birth (i.e. 4th month on the file), or 6 months after birth, or 9 months after birth, etc.

It should be mentioned that by collapsing cycles into groups of three and by selecting a portion of the individual monthly age segments (after birth on the BRMF) that some 'smoothing' has been generated (i.e. the impact of any outliers has been reduced).

The counts for these active unclassified records are given in Table 8. Also indicated are

the proportions (in %) of each set-of-cycle's births that are active 70099's for each quarter after their birth. Table 8.A shows the average proportion (in %) of births across sets-of-cycles that are active 70099's for each quarter where the births are limited to those cycles for which counts for the quarter are available. (An indication of how the birth counts were grouped is also given in Table 8.A).

The following observations can be made from Tables 8 and 8.A:

i) On average, 26.13% of births are classified by three months after being added to the BRMF; an additional 38.29% are classified within the next 3 months.

ii) The median time lag between being added to the BRMF and being classified is between 3 and 6 months. (The first quartile is approximately 3 months, and the third quartile is approximately 9 months.)

iii) An average of 18.68% of births (or 7,571 every 3 months) are unclassified one year after having been birthed on the BRMF. After an additional year, 3.49% still remain unclassified.

The proportions given in Table 8 have been plotted in Graph 1; the proportions for each quarter after birth have been joined, and the average proportion for each from Table 8.A has been drawn as a broken horizontal line.

A recent investigation called the PD Processing Time Lag Study [3] produced estimates of time lags between specific events for various categories of PD account births on the BRMF in March 1984. Two of those time lags which are relevant here are

a) PD20 data capture to birth of record on the BRMF and

b) PD20 data capture to classification of record on the BRMF.

The quartiles for these two time lags are given in Table 9; they relate to that subset of the PD account births which were identified as being new accounts for new businesses which were classified within 8 months after being birthed.

The counts given in Table 8 can also be used to evaluate the effectiveness of the follow-up program. Instead of calculating proportions based on the number of records birthed, proportions could be calculated using the number of unclassified active records as of the fourth quarter after birth as the base/denominator. This is, after all, the number of records to which the follow-up procedures are applied.

At this point it might be appropriate to mention what activities have an impact on the counts of active 70099 records. The number of such records decrease if there is classification (resulting from PD20 processing or follow-up) or inactivation (from follow-up). The counts may even increase if there is reactivation of accounts.

The proportions based on the number of active 70099 records as of the fourth quarter after birth are given in Table 10. Again portions of the total number of records (for follow-up) were used to calculate the average proportions for each quarter after follow-up, as indicated in Table 10.A.

From these average proportions it can be seen that the median time lag between being identified for follow-up (at the fourth quarter) to classification is 3 to 6 months (i.e. between the fifth and 6th quarters after birth). After one year of follow-up (i.e. as of the eight quarter after birth), almost 20% of records to be followed up have not yet been classified.

These post-follow-up proportions and their averages have been plotted (see Graph 2.).

Table 8

Number of Active 70099 Records for Each Quarter After Birth

Set of BRMF Cycles	Number of Records Birthed	Quarter After Birth								
		1	2	3	4	5	6	7	8	9
157-159	33,088									
154-156	42,270	31,849 75.35%								
151-153	53,946	43,342 80.34%	20,854 38.66%							
148-150	47,216	35,648 75.50%	16,252 34.42%	12,273 25.99%						
145-147	33,658	26,549 78.88%	10,263 30.49%	6,773 20.12%	4,811 14.29%					
142-144	46,344	38,238 82.51%	17,623 38.03%	11,120 23.99%	7,099 15.32%	4,067 8.78%				
139-141	54,526	50,385 92.41%	25,508 46.78%	18,994 34.83%	13,960 25.60%	6,178 11.33%	2,227 4.08%			
136-138	45,131	29,117 64.52%	14,300 31.69%	11,125 24.65%	8,323 18.44%	4,692 10.40%	3,643 8.07%	2,239 4.96%		
133-135	28,647	14,916 52.07%	8,833 30.83%	7,176 25.05%	5,558 19.40%	4,171 14.56%	2,152 7.51%	2,013 7.34%	1,527 5.33%	
130-132	34,814	15,516 44.57%	8,861 25.45%	7,492 21.52%	5,673 16.30%	5,390 15.48%	1,813 5.21%	703 2.02%	687 1.97%	470 1.35%

Note: the percentages are the proportions of each set-of-cycle's births that are active 70099's.

Table 8A

Calculation of Average Proportions for Active 70099's by Quarter After Birth

Sub-totals of the number of records birthed

Set of BRMF Cycles	Number of Records Birthed	Quarter After Birth								
		1	2	3	4	5	6	7	8	9
157-159	33,088									
154-156	42,270	386,552								
151-153	53,946		344,282							
148-150	47,216			290,336						
145-147	33,658				243,120					
142-144	46,344					209,462				
139-141	54,526						163,118			
136-138	45,131							108,592		
133-135	28,647								63,461	
130-132	34,814									34,814
Number of Active 70099 Records		285,560	122,494	74,953	45,424	24,498	9,835	5,045	2,214	470
Average Proportion of Births that are Active 70099 Records		73.87%	35.58%	25.82%	18.68%	11.70%	6.03%	4.65%	3.49%	1.35%

Graph 1

Proportions of Births That are Active 70099 Records by Set of BRMF Cycles for Each Quarter After Birth

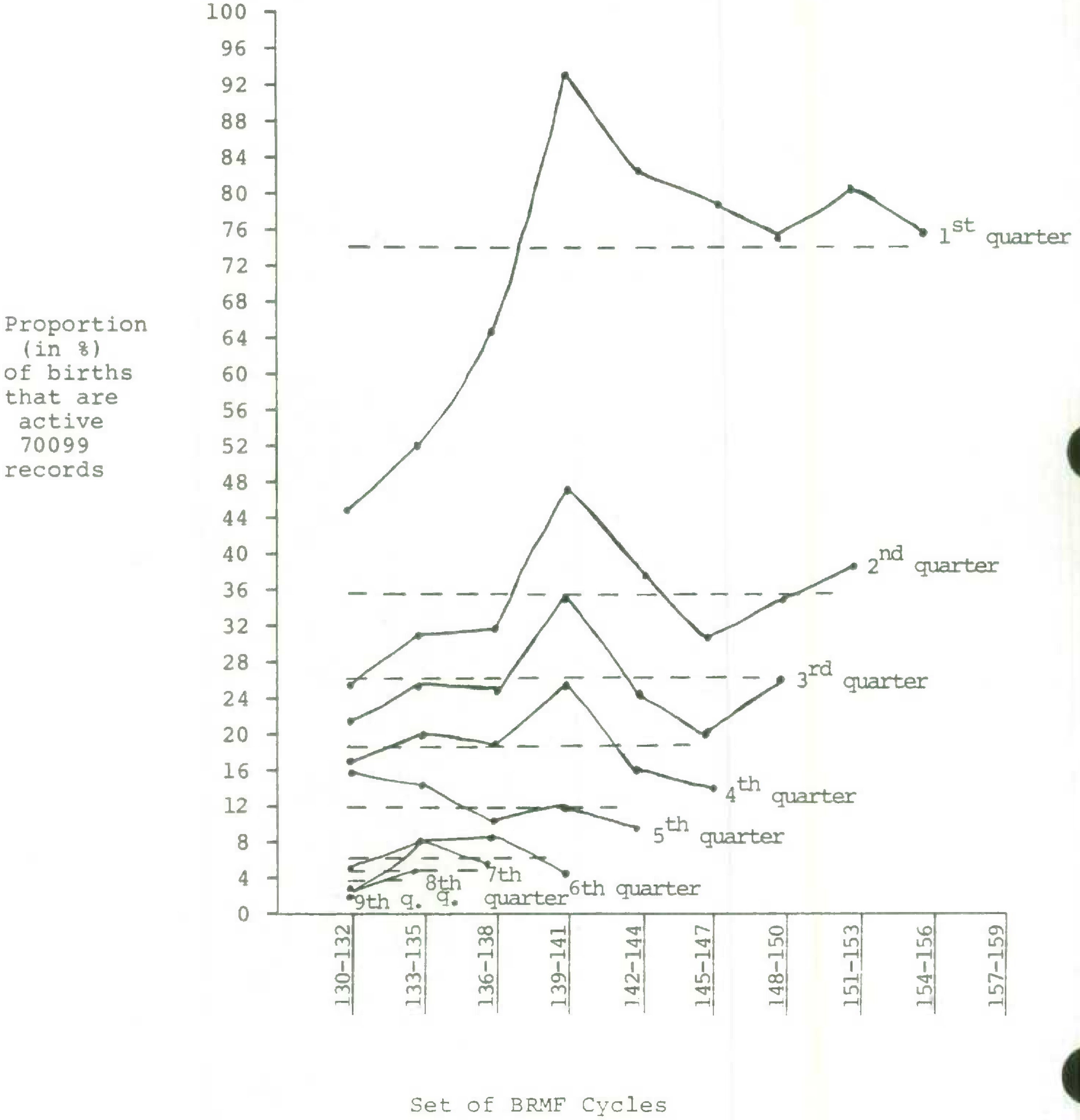


Table 9

Estimates of the Quartiles for the Durations for PD20 Related Time Lags for PD Account Births* for New Businesses

Quartile	PD20 data capture to birth on BRMF	PD20 data capture to classification on BRMF
1	-1 to 0	2 to 3
2	0 to 1	4
3	0 to 1	4 to 5

* these estimates are based solely on a subset of the PD accounts which were birthed on the BRMF as of March 13, 1984.

Table 10

Number of Active 70099 Records for Each Quarter After Follow-up and Their Proportion of the Number of Active 70099 Records to be Followed-up for Each Set of BRMF Cycles

Set of BRMF Cycles*	Number of Records to be Followed-up	Quarter After Follow-up				
		1	2	3	4	5
142-144	7,099	4,067 57.29%				
139-141	13,960	6,178 44.26%	2,227 15.95%			
136-138	8,323	4,692 56.37%	3,643 43.77%	2,239 26.90%		
133-135	5,558	4,171 75.04%	2,152 38.72%	2,103 37.84%	1,527 27.47%	
130-132	5,673	5,390 95.01%	1,813 31.96%	703 12.39%	687 12.11%	470 8.28%

* these cycle numbers refer to the BRMF on which these records were birthed, and not to the BRMF on which the follow-up was based.

Table 10A

Calculation of Average Proportions for Active 70099 Records

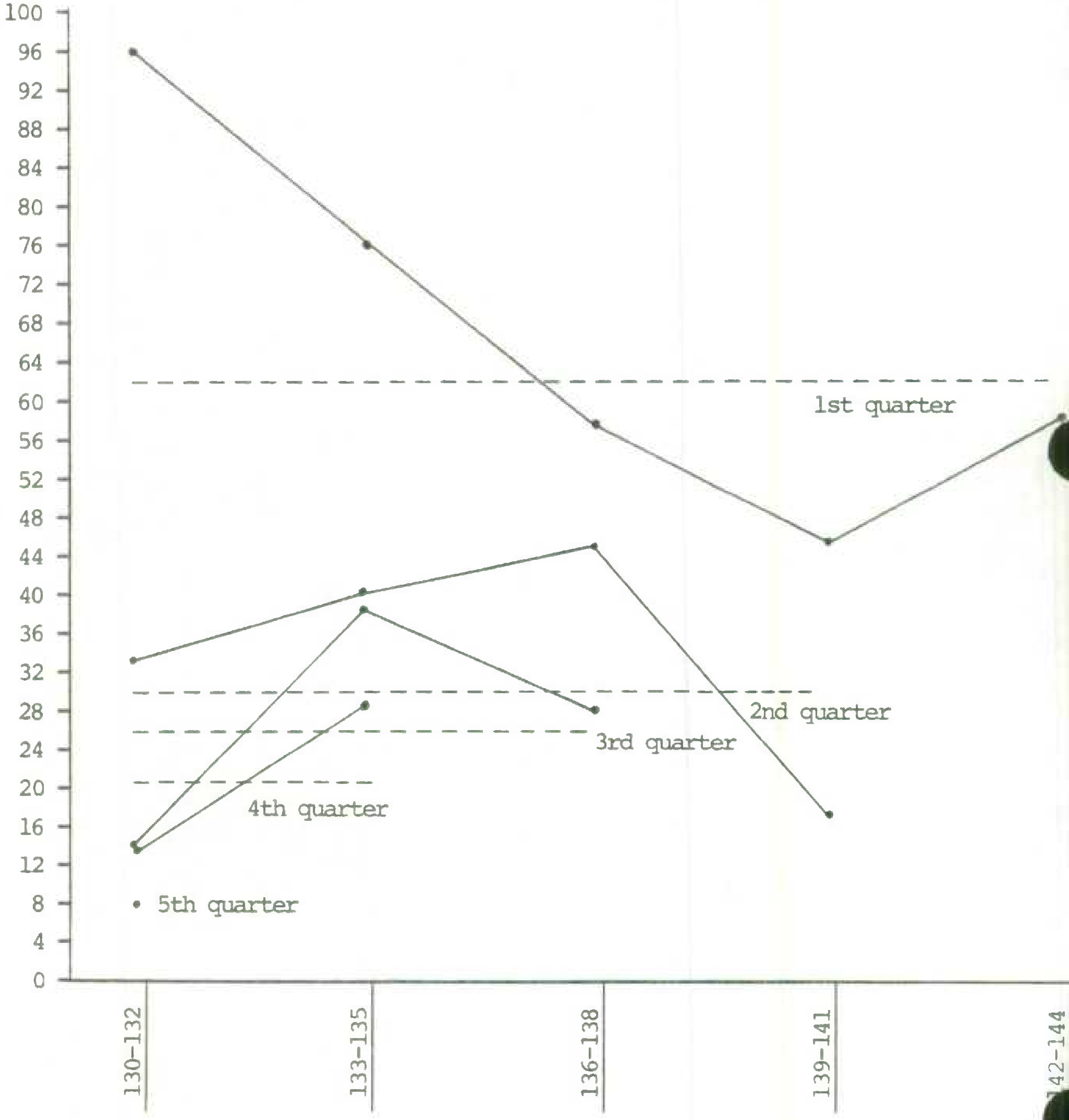
by Quarter After Follow-up

Set of BRMF Cycles*	Number of Records to be Followed-Up	Quarter After Follow-up				
		1	2	3	4	5
142-144	7,099	40,613		Sub-totals of the number of records to be followed-up		
139-141	13,960		33,514			
136-138	8,323			19,554		
133-135	5,558				11,231	
130-132	5,673					5,673
Number of active 70099 records		24,498	9,835	5,045	2,214	470
Average proportion of records which are active 70099's		60.32%	29.35%	25.80%	19.71%	8.28%

[*] these cycles refer to the BRMF on which these records were birthed, and not to the BRMF on which the follow-up was based.

Graph 2

Proportions of Records to be Followed-up
that are Active 70099,s
by Set of BRMF Cycles
for Each Quarter After Follow-up



As counts were available for only a limited number of sets of BRMF cycles, data on the age distribution of active 70099's for each set of BRMF cycles were also limited. Table 11 shows the number of active unclassified records for the first four quarters after birth on the BRMF for each of 6 recent sets of BRMF cycles. Based on the total of counts for each set of cycles, proportions for each quarter were calculated. These proportions were plotted against the sets of BRMF cycles in Graph 3; the points for each quarter after birth have been joined.

From these data, the following observations have been noted:

i) For every set of cycles, the 1st quarter after birth is always a higher proportion (of the total) than that for any other quarter after birth.

ii) Only 3 sets of cycles (i.e.

142 - 144 }
145 - 147 } "old"
and 157 - 59 - the most recent)

have every-decreasing proportions for increasing age (i.e. an increasing number of quarters after birth). Presumably this is a desirable characteristic. At the other extreme, for cycles 151-153 the "oldest" 70099's (i.e. from the 4th quarter after birth) were more prevalent than the 70099's in the second youngest (2nd quarter after birth) group.

It should be noted, however, that births to the BRMF are seasonal and the amount of human resources available to process PD20 and follow-up documents are also seasonal, so that these kinds of variations in the age distribution of 70099's within each BRMF may be the result.

Table 11

Number of Active 70099 Records for Each Set of BRMF Cycles*

by Quarter After Birth

Set of BRMF Cycles*	Quarter After Birth				Total (see Note)
	1	2	3	4	
157-159	31,849 45.64%	20,854 29.88%	12,273 17.59%	4,811 6.89%	69,787 100.00%
154-156	43,342 59.00%	16,252 22.12%	6,773 9.22%	7,099 9.66%	73,466 100.00%
151-153	35,648 50.21%	10,263 14.46%	11,120 15.66%	13,960 19.66%	70,991 100.00%
148-150	26,549 37.14%	17,623 24.65%	18,994 26.57%	8,323 11.64%	71,489 100.00%
145-147	38,238 47.54%	25,508 31.71%	11,125 13.83%	5,558 6.91%	80,429 100.00%
142-144	50,385 64.98%	14,300 18.44%	7,176 9.26%	5,673 7.32%	77,534 100.00%

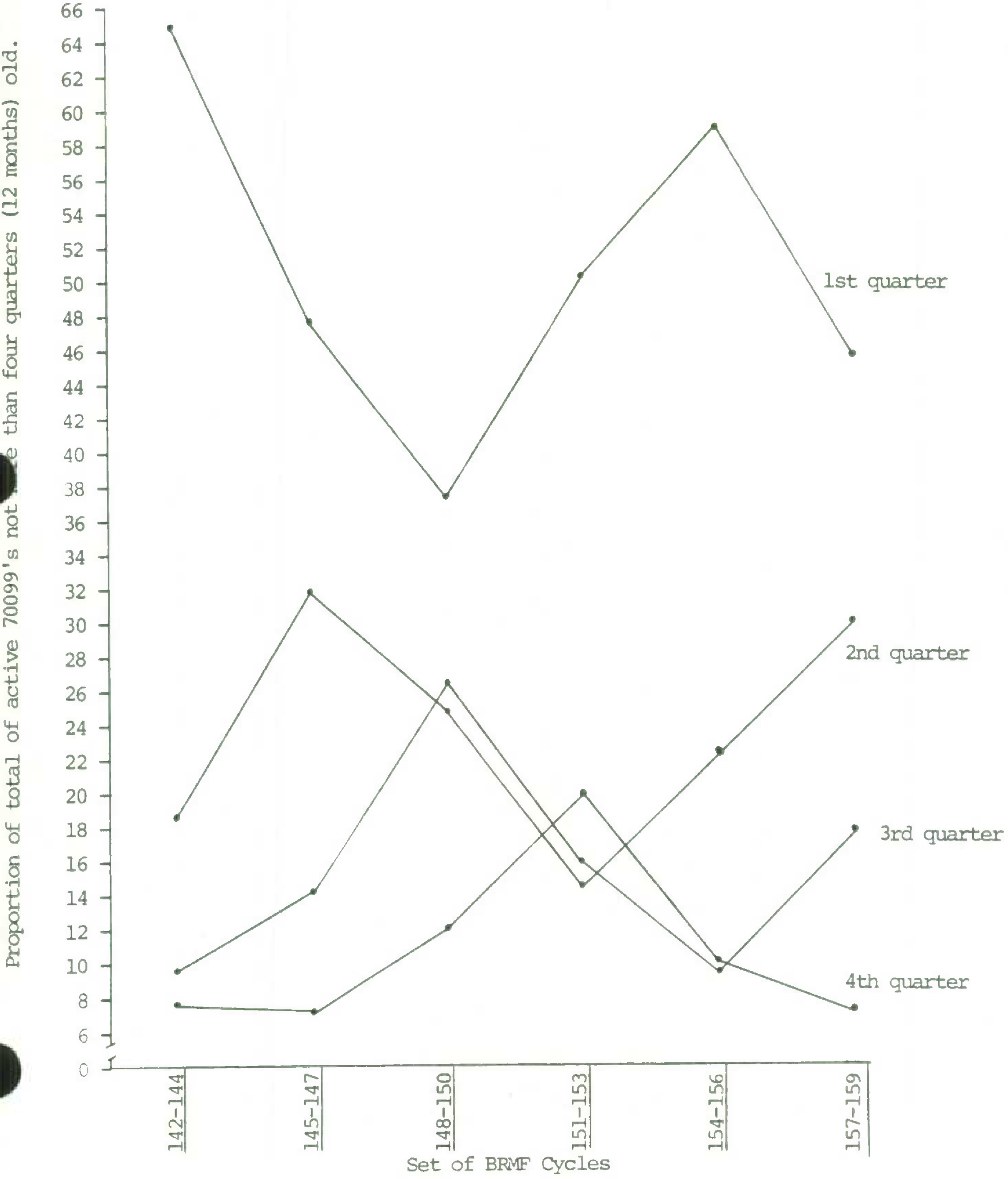
* These cycle numbers refer to the BRMF's on which these records were identified as being active 70099's, and not to the BRMF's on which they were birthed.

Note: The total is the number of active 70099 records that are no more than 4 quarters (12 months) old.
The percentages are the proportion for each quarter of the total for each set of cycles.

Graph 3

Proportion of Active 70099,s for Each Set of BRMF Cycles

By Quarter After Birth



3.2.3 Recommended Future Work

i) It is recommended that investigation be carried out for unclassified PD-defined records by monitoring them on the monthly PAYDAC files received in STC from RC-T; specifically a cross-classification of Current Year Balance (CYB) by Registration Status Code (based on Account Type and Activity Status) would give an indication of the undercoverage of surveys due to 70099's.

ii) Because births to the employer universe are seasonal, and as the resources available to process updates to the frame are also seasonal, there may, as a result, be some seasonal variation in the quality of the frame. The implications of such a situation should be reviewed.

iii) It is recommended that the current follow-up strategy for unclassified records be reviewed with the perspective of having a priority scheme for following-up and classifying these records. This review should include the resource cost of each new proposal.

iv) Consideration should be given to executing the follow-up for unclassified records before they are twelve months old.

v) The effect of birthing BRMF records from PD20 forms should be studied. This would be limited to those accounts whose PD20 is received in STC before the PD account number is observed on PAYDAC.

v) The effect of birthing BRMF records from PD20 forms should be studied. This would be limited to those accounts whose PD20 is received in STC before the PD account number is observed on PAYDAC.

vi) The current procedures for classification and unduplication (including document handling) should be reviewed with the perspective of reducing the time lag between birth and classification.

4. Conclusion

This report has looked at the past and current use of the PD data (i.e. both the PAYDAC and PD20) and has shown many deficiencies in providing frame data to the current BRMF. One should therefore not assume that the current PD processing should be transposed as is into the CFDB environment. The use of the T1 and T2 data as well as the potential addition of an area frame may modify the current objectives STC has in acquiring the PD data. Even if the objectives were kept the same, enhancements in the processing must be made to resolve the known deficiencies.

Therefore, future evaluations made by Team 1 should be related to the future usage of the PD data in the CFDB environment. Thus, the objective(s) that the CFDB has in acquiring the PD data should first be clearly spelled out. Then the future evaluation work, especially the one listed in this document, should be prioritized and organized into a framework that will assess how well the PD data can fulfill the objective(s).

5. References

[1] Bankier, M. (1985), "Methodology Involvement with Sub-Project Team 1 of the Infrastructure Project", BSMD Memo, March 13, 1985.

[2] Infrastructure Sub-Project Team 1, "Preliminary Report on Documentation of Administrative Data Sources", May 29, 1985.

[3] Beckstead, D. (1985), "PD Processing Time Lag Study - The Final Estimates - A Methodological Report", BSMD, April 1985.

[4] Employment Statistics Development Project, "The Report of the PD7 Analysis Team", July 1979.

[5] Estevao, V. (1985), "A Report on the Quality of the Data in the BRMF, SARUS Study - 1984/5", May 2, 1985.

[6] Finlay, F. (1982), "A Report on the Identification of Employee Status on the BRMF: A Proposal for Identifying Non-employers", BRD, December 1982 **(DRAFT)**

[7] Estevao, V. (1983), "A Review of the Report on the Identification of Employee Status on the BRMF", BSMD, February 16, 1983.

APPENDIX A

1970 S.I.C. Codes Defining Industrial Activity Not of Interest to Surveys of Economic Production

001, 003, 011, 013, 015, 017, 019 021	Agriculture
801-807, 809	Education and Related Services
821, 822, 828	Hospitals, Related Health Care Institutions and Welfare Organizations
831	Religious Organizations
873	Private Households
902, 909, 931, 951, 991	Federal Administration