HEALTH ECONOMICS RESEARCH, INC.

300 Fifth Avenue, 6th Floor
Waltham, MA 02154
(617) 487-0200
(617) 487-0202 Fax

UNIQUE PHYSICIAN IDENTIFICATION NUMBER (UPIN) VALIDATION STUDIES

Carrier Analysis

Submitted by:

Killard W. Adamache, Ph.D.
Holly A. Cyr, B.A.
Angela R. Merrill, B.A.

Prepared for:

Ann Meadow, Sc.D.
HCFA Project Officer

August 23, 1994

This report was prepared under Delivery Order No. 5 of the HCFA Master Contract No. 500-92-0020. The views and opinions expressed in this report are the authors and no endorsement by HCFA is intended or should be inferred.

UNIQUE PHYSICIAN IDENTIFICATION NUMBER (UPIN) VALIDATION STUDIES

PREFACE

The *Physician Registry* was established in 1989 by HCFA for the purpose of issuing unique identification numbers (UPINs) to the following Medicare providers – doctors of medicine, osteopathy, dental medicine, dental surgery, podiatry and optometry, and chiropractors. To assist in the identification of individual providers, the *Registry* collects a variety of information. HCFA contracted with Health Economics Research, Inc. (HER) to perform studies on the *Registry's* data quality and on the consistency of physician information on Medicare claims. Another objective was to develop data edits both to assist carriers in maintaining their *Registry* data and to assist researchers in using *Registry* data.

The current report is the first of four reports produced under the HER contract. It describes the results of an analysis of the quality of *Registry* data. It also provides estimates of the number of providers that might have more than one UPIN. The second report provides data edits for use by carriers in maintaining their *Registry* data. The third report provides data edits for use by researchers using *Registry* data. The fourth report has several components, including: (1) estimates of the number of HMO physicians without UPINs who provide services to Medicare beneficiaries, (2) an analysis to identify the inadvertent issuance of UPINs to group practices and corporations, (3) an analysis of the consistency of a physician's Medicare participation status on the *Registry* and on the physician's associated Medicare claims and (4) a case study of the operations of the *Registry* contractor and carrier operations. It also contains an executive summary of the results for all four deliverables. An additional deliverable consists of documentation for SAS files submitted to HCFA. The SAS files contain special variables that indicate the quality of data elements maintained by the *Registry*, the consistency between pairs of data elements, and an indication whether a provider might have more than one UPIN.

TABLE OF CONTENTS

PAGE

TABLE OF CONTENTS (continued)

PAGE

TABLE OF CONTENTS (continued)

PAGE

TABLE OF EXHIBITS

TABLE OF TABLES

TABLE OF TABLES (continued)

TABLE OF FIGURES

1.0 BACKGROUND AND PROCEDURES OF THE PHYSICIAN REGISTRY

Congress saw a need for Medicare to establish a unique identifier to support administrative functions such as uncovering fraud and abuse. In response to Congress, a system was established to issue unique identifiers to medical doctors and similar professionals. This chapter provides an overview of the establishment and operations of the Physician Registry, the administrative and research uses of the unique identifiers issued by the Registry, and the use of the identifiers on Part B claims submitted to HCFA. The material in this chapter is drawn primarily from the *Medicare Carriers Manual Transmittals* 1 through 7 (Part 4, Professional Relations) from April 1989 through May 1993 and from *Program Memorandum to Carriers* (November 1991 and March 1992). As such, then, the overview presented in this chapter can be construed as a theoretical discussion of Registry operations.

1.1 Introduction

The Consolidated Omnibus Budget Reconciliation Act (COBRA) of 1985 (P.L. 99-272), section 9202, required HCFA to assign a unique identifier to all physicians to which payment is made under Medicare. The resulting master file of Unique Physician Identification Numbers (UPINs), maintained under the Medicare Physician Identification and Eligibility System (MPIES), is the first national database of individual Medicare physicians.

As of July 1993, over 693,355 UPINs had been assigned to all Medicare physicians[1] including doctors of medicine, osteopathy and podiatric medicine, doctors of optometry, doctors of dental surgery and medicine, and chiropractors. While a separate record is kept for

[1]The Social Security Act, Section 1861(r), defines a doctor of medicine, doctor of osteopathy, doctor of dental surgery and dental medicine, doctor of podiatry, doctor of optometry, and chiropractors as physicians for purposes of Medicare payment under Title 18 Part B.

each physician's practice settings, only one UPIN is assigned. This UPIN will remain associated with the physician throughout their affiliation with Medicare. HCFA has contracted with Transamerica Occidental Life Insurance Company (TOLIC) to establish the MPIES, or the Registry, which began operations in January 1989.

The Omnibus Budget Reconciliation Act (OBRA) of 1989 required the reporting of the referring/ordering physician's name and UPIN on Medicare claims for selected services for which a referral was given. OBRA90 required HCFA to publish a national directory of UPINs to accommodate the referring UPIN requirement. Compliance with the reporting of UPINs on Medicare claims, and the validity of the UPINs reported, is important for future Medicare Part B program management, policy, and research using the National Claims History (NCH) and for administrative purposes as well.

This report presents the results of an analysis of data maintained in the Registry. The study was prompted by problems found in the data by HCFA researchers and their contractors. This report consists of nine chapters. The first chapter provides an overview of the entire Registry system, including the specified duties of TOLIC and the carriers. It also reviews the importance of UPINs for claims processing, policy research, and enforcement of HCFA regulations. From site visits conducted during August and September of 1993, the second chapter summarizes TOLIC and carrier activities that affect the quality of data in the Registry database. Criteria for classifying data element values are discussed in the third chapter. Data element integrity measures are presented and analyzed in the fourth chapter. The logical consistency between pairs of data elements is the subject of the fifth chapter. Inconsistent values of data elements across provider practice settings is the subject of the sixth chapter. Inconsistencies between the header and practice setting versions of data elements are examined in the seventh chapter. Estimates of the number of providers with more than one UPIN are presented in the eighth chapter. A summary of the empirical results is presented in the ninth chapter.

1.1.1 Need For Unique Physician Identifier Numbers

1.1.1.1 Research Purposes

The use of UPINs by physicians billing or referring services paid under Medicare allows for more accurate and complete reporting of individual physician utilization of services and referral patterns. NCH claims data will contain information on individual physicians rather than group practices, allowing researchers to disaggregate group practice utilization and payment information to the physician level. Also, a UPIN allows aggregation of information on a physician who submits claims to more than one carrier or who, within a single carrier jurisdiction, practices in more than one setting. The presence of the referring/ordering physician's UPIN on a claim enables researchers to link consultations and ancillary services, such as laboratory tests, with initial visits to a physician; such linkage is useful when investigating alternative payment policies, such as payment bundles for Part B services.

The billing of Medicare services by interns and residents may also be traceable by UPIN in the future. Interns and residents began receiving UPINs in 1992, and beforehand were identified by a surrogate UPIN, as discussed in section 1.2.2.3.

1.1.1.2 Administrative Purposes

UPINs provide important administrative benefits such as improving accuracy in claims processing and record keeping, as well as aiding HCFA to enact payment safeguards. The presence of a UPIN also makes it easier for HCFA to identify sanctioned physicians, as all sanction information on a physician should be on the MPIES record(s).

1.2 Background

1.2.1 UPIN Assignment Process

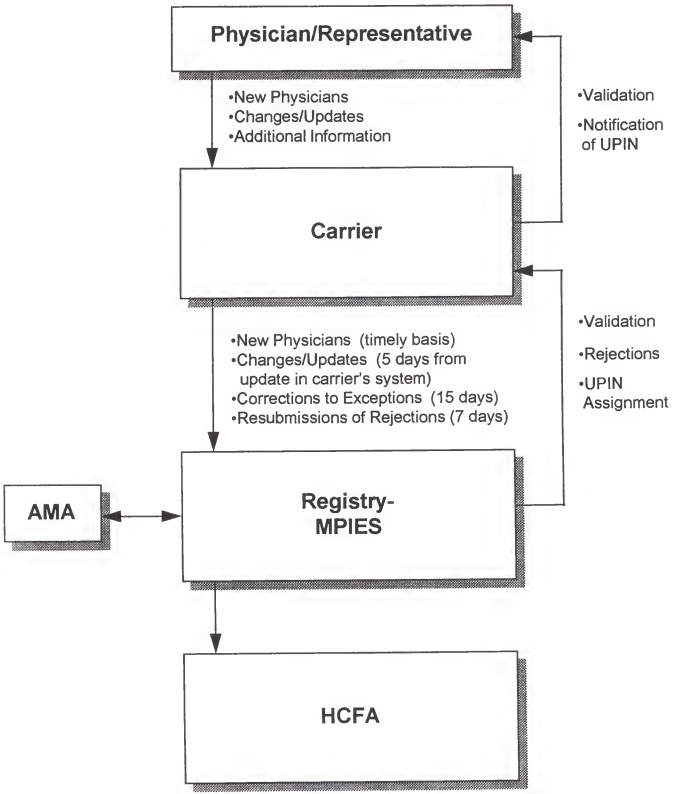
Before the assignment of a UPIN, the carriers must first collect and validate information from the physicians (or their representatives) and then submit this information to the Registry for further validation/edits. Exhibit 1-1 illustrates the process of obtaining a UPIN from the Registry and maintenance of the MPIES. The carriers receive periodic instructions on collecting and validating physician information and on the assignment of UPINs through transmittals in the Medicare Carriers Manual (MCM) Part IV Professional Relations and through both Part A/B and Part B Program Memoranda from HCFA.

Before explaining in detail the process depicted in Exhibit 1-1, a brief overview is provided of the day-to-day process of UPIN assignment and maintenance. Whether making an initial application to receive a UPIN or adding a new practice setting, a provider (or representative) needs to submit an application to the local Medicare carrier. The carrier then processes the application. If required data elements (e.g., date of birth) are not present on the application or are not available in the carrier's provider file, then the carrier is supposed to obtain the requisite information from the provider before submitting the provider's information to TOLIC.

Once the requisite provider information has been obtained, the carrier submits the information in a standard format ("carrier record") to TOLIC. If the practice setting is to be attached to a provider that had already obtained a UPIN and the carrier had included it on the carrier record, TOLIC takes the new information and adds a new practice setting segment to the provider's master Registry record. Before doing so, however, TOLIC checks the provider name. If the provider name differs, then TOLIC validates, if necessary, the change with the carrier.

Exhibit 1-1

UPIN ASSIGNMENT PROCESS AND MAINTENANCE OF THE MEDICARE PHYSICIAN IDENTIFICATION AND ELIGIBILITY SYSTEM (MPIES)



If an incoming carrier record to the Registry does not have a UPIN on it, TOLIC, using its record matching procedures, determines whether the incoming carrier record is for a provider already in the Registry or whether a new UPIN needs to be issued. If necessary, under a sub-contract, the American Medical Association (AMA) is consulted during the provider identification process. When the Registry issues a UPIN for a provider, TOLIC sends a notice to the carrier. It is then the carrier's responsibility to send a letter, with the new UPIN, to the provider.

1.2.1.1 Physician Application and Carrier Validation

Physician Application Requirements

All physicians who are billing for Medicare services must obtain a UPIN in order to be reimbursed by HCFA. To help inform physicians of the necessity of obtaining a UPIN, carriers are expected to contact physicians directly or through state and local medical societies and trade associations with carrier bulletins, newsletters, and other mailings.

To receive a UPIN physicians must submit all required data elements for validation to the carrier. After ascertaining that all data are correct, the carrier will then submit the physician record to the Registry for the assignment of a UPIN. The carrier must maintain individual physician enrollment records regardless of a physician's affiliation with a group or clinic. For each physician practice setting, the physician currently must submit the following "minimum set" of data elements[2]:

- Professional name;
- Addresses (business and billing);

[2]Until December 1991, state license number, medical school, and graduation were not mandatory data elements.

- State in which physician is licensed in a specific practice setting;
- State license number in a specific practice setting;
- Date of birth;
- Credentials (MD, DO, CH, etc.);
- Graduate medical school;
- Year of medical school graduation;
- Specialty;
- Resident/Intern status;
- Previous practice in another state;
- Indicator for group practice members; and
- Physician/Group Medicare participation status.

Carrier Validation

The carrier must then conduct the validation process which includes verifying state licensure and other information with the appropriate state licensing board; verifying physician status for prior practices with the carrier serving the prior practice (including collecting prior utilization information in order to determine whether the physician should be placed on pre-pay review for aberrant practices); and verifying that the physician has not been sanctioned by checking the Medicare/Medicaid Sanction-Reinstatement Report.

To collect and maintain data on physicians within group practices, the carrier is expected to survey the groups annually to update the listing of the members of the group. In addition, the groups must report to the carrier any changes in their membership as they occur. The minimum data elements described above must be collected for each new addition.

1.2.1.2 UPIN Carrier Record

Carrier Record Description

In order to standardize the provider information submitted by the 59 carriers to TOLIC, a standardized carrier record was developed. It specifies the data elements to be included, the order of the date elements, and the format for each data element.

The UPIN carrier record consists of 41 data elements, 34 of which are used in the assignment of a UPIN to the physician by the Registry. The remaining seven elements (UPIN, Transamerica number, Registry assigned error codes/notification codes, Record validation field, Special processing indicator, Special processing data, and a Filler) are entered by the Registry and may be used later in the validation/edit process between the Registry and the carrier. The first data element on the carrier record is the Record Code, which identifies the reason for transmitting the new record (either to or from the carrier). Reasons include: adding a new physician or practice setting, correcting a previously submitted record, updating a record, transmitting a newly assigned UPIN, *et cetera*. This data element is important because it identifies the type of record to the carrier and the Registry, who during the validation/edit process may be passing a record back and forth. All data elements on the carrier record are retained as MPIES elements except for the physician's business and billing street, city, and state addresses.

Carrier Record Submission

Once all physician information has been collected and validated by the carrier, the UPIN carrier record for each physician and practice setting is submitted to the Registry. The carrier may submit records to the Registry by telecommunications or by mailing diskette/tape. In order to be accepted by the Registry, the UPIN carrier record must adhere to strict Registry

requirements or it will be rejected and returned to the Carrier. For example, if a record is submitted on disk or tape it must have the correct external label.

Records submitted on tape or diskette will be processed within 24 hours of receipt. If the submission is not accepted by the Registry for reasons such as incorrect format or unreadable density, the entire file will be rejected. The file should then be corrected and resubmitted within ten working days or two calendar weeks, identified as a resubmittal.

Unless the volume on a daily or weekly basis requires excessive transmission time, the carriers are encouraged to submit the carrier records electronically through the telecommunications network that exists between the Registry, the carriers, and the AMA. Records submitted electronically will be processed by the Registry's system on the same day received.

1.2.1.3 Additions and Updates to MPIES

To ensure the integrity of the information in the Registry, carriers are asked to perform routine maintenance on their physician file and promptly submit notification (i.e., the UPIN carrier record) to the Registry of any adds, corrections of exceptions, and updates. Each time an add, correction, or update is submitted, the entire carrier record for that physician as discussed in section 1.2.1.1 is included. "Adds" refer to the submission of any new physician who needs a UPIN, any new practice settings of a previously registered physician, and any changes in group membership. Updates consist of the revision of any critical data elements on a record of a previously registered physician. Exceptions are corrections and resubmission of records that were returned to the carrier for development. These three types of maintenance records may be submitted daily; however, no maintenance record should be older than five days from the date of update in the carrier's system.

Adds

When submitting an add, the carrier must submit the entire carrier record with current information in each data element. An add should have a Record Code "1" (add) in the first data element. All data elements on the carrier record (except the street and city address) are retained on the MPIES and are used, when necessary, to resolve questions regarding the identity of a physician.

For add records with a UPIN, the provider's last name and the first letter of the first name on the add record need to match the corresponding values on existing the UPIN record on the Registry master file. If a match is found and the data elements on the add record satisfy the data screens, then a new practice setting is added to the provider's record.

For add records with a UPIN that do not have the requisite name match and for carrier records submitted without a UPIN, TOLIC checks provider records in the master file until a match is found. If a computerized match is found and all of the values of selected data elements on the add record are equal to their counterparts on the "matched" record on the master file, then a new practice setting is attached to provider record.

If a match is not found on the Registry master file, TOLIC then checks an AMA data extract. The physician's name, address, date of birth, and state licensed are used for matching against the AMA data. If a computerized search of the AMA data extract does not result in a match, TOLIC's system will flag the add record for manual review. A TOLIC analyst then reviews existing UPIN records with the provider's last name providing the basis for the initial screening element. When reviewing records, the analyst has full access to both header and practice setting information. If a match to an existing Registry record still cannot be found, a new UPIN is issued by TOLIC.

Updates

Carriers will have updates to existing records for "critical" physician information. Updates can be triggered by changes in zipcode, change in group practice, death, change in Medicare Participating Physician program status, *et cetera*. With the lag between physician notification of the carrier and carrier notification of the Registry it is possible to have two different records on the same physician existing at one time (one at the carrier and one at the Registry). Therefore, a timely update process is necessary.

When a carrier submits an update to the Registry, it sends the entire carrier record with revised information in the correct elements and a Record Code "5" (MPIES update) in field 1. As this physician will have previously been assigned a UPIN, the update record should also show the UPIN in field 34, the carrier number in field 36, and the physician's provider number in field 33. The Registry will then validate the updated information for feasibility.[3] If the UPIN, carrier number, and provider number of the update record and of the existing MPIES record do not match exactly, the record will be rejected and returned to the carrier. Once the Registry assigns a UPIN, that UPIN, the carrier number, and the provider number can only be changed using a special update process because these three fields are essential in ensuring that the correct record is being updated. Name changes require special processing because they also affect key matching fields. An update that is submitted for a name change should include a Record Code "5," the physician's updated name information in fields 3, 4, 5, and 6, the provider number, the UPIN, the carrier number, a Special Processing Indicator of "1" in field 39, and the physician's old name in field 40. The name in field 40 must match exactly the existing MPIES record or the update will be returned. Similarly, the change of a provider number requires a special process as the provider number on the existing MPIES record is a key matching field. An update record for a change in provider number will have a Special

[3]The Registry determines whether a value is feasible (e.g., valid specialty code) but cannot determine whether a feasible value is accurate (e.g., whether the specialty code should be equal to 10 or 11).

Processing Indicator of "2" and the provider's old number in field 40. Also, an update to change the provider number must be submitted separately from any other updates.

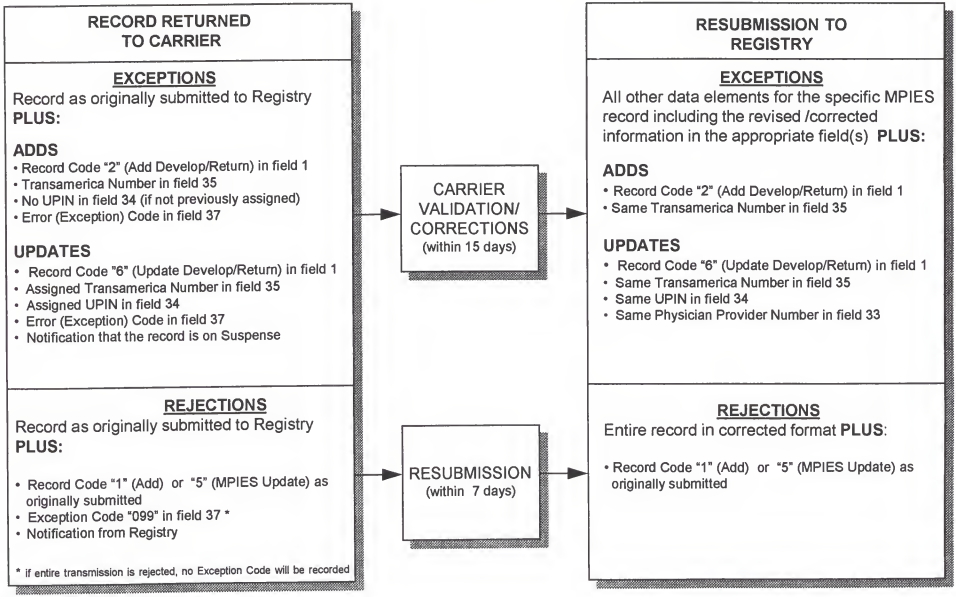
Exceptions

Any record (add or update) that contains data in all the required fields, but cannot enter the MPIES for any reason is an "exception." Exceptions are held in suspense until all corrections are made and a UPIN is assigned (in the case of adds involving assignment). When a record is held in suspense, the Registry returns a copy of the record to the carrier with error codes indicating that the record is an exception and showing which fields need editing or further validation. If the exception is an add, the Registry will not add that record to the MPIES until all discrepancies have been resolved. Exhibit 1-2 illustrates the exception process. Each exception record is returned to the carrier and may include up to five three-digit codes in field 37- Registry Assigned Error Code. The first digit is used for internal processing validation while the last two digits are the Exception Code (e.g., "last name"). Up to five data elements needing validation/correction will be indicated in field 37; if more than five elements exist that require missing information or verification, the field will contain "099" indicating that the entire record should be validated.

The carrier must then validate all elements indicated in field 37, or the entire record, before resubmitting the record to the Registry. The carrier must first check that the information submitted was accurate according to their in-house files, both automated and hard copy files, such as physician applications. Next the information should be again validated with the State Licensing Board. The carrier can also check with the local professional association; however, no missing information should be taken from any copyrighted source such as the *American Medical Directory*. The carrier may also need to recontact the physician. If a physician is contacted, the carrier should verify the entire set of data elements on the record, not just the ones on which there are questions.

Exhibit 1-2

IDENTIFICATION AND RESUBMITTAL OF REGISTRY EXCEPTIONS AND REJECTIONS



Exceptions should be submitted as they are corrected, rather than waiting until a whole batch is corrected, to minimize the number of records the Registry holds in suspense. Exception records should be corrected and resubmitted within 15 calendar days.

Rejections

The Registry will reject records for technical reasons such as incorrect tape density or label and also for reasons such as an incorrect Record Code, wrong or missing carrier ID number, invalid physician status code, or an alpha character in a numeric field. When a batch of records is being submitted, the entire transmission will be rejected if a transmission format violation exists (e.g., no trailer record), if the trailer record count does not equal the number of records in the transmission, or if more than ten percent of the records have any of the errors listed above. If less than ten percent of records have errors, the batch will be accepted but the individual erroneous records will be returned to the carrier for corrections. The Registry is responsible for telephoning the carrier and following up by letter when any records are rejected. The Record Code of a returned record will remain "1" or "5" as originally submitted while the Exception Code "099" will be recorded in field 37 of all rejected records. The carrier should correct and re-submit rejected records within seven calendar days.

1.2.1.4 Assignment of UPIN

Once the Registry has assigned a UPIN, a physician record is placed on the MPIES. This record is also sent back to the carrier which submitted it, who must then notify the physician of the assigned UPIN. There is no special mailing required to issue the UPIN to the physician; however, notification must be mailed within 30 days of receipt of the UPIN from the Registry. The carrier sends the UPIN to each of the physician's practice settings although only one UPIN is assigned.

The notification letter to the physician, especially to physicians in a group practice, should be marked "Personal and Confidential" and must contain a standard message about the UPIN (as published in an October 1989 transmittal to the carriers). Physicians are advised that they may receive more than one notification if they have different practice settings, but to call the carrier immediately if these notifications indicate more than one UPIN. According to a March 1992 Program Memorandum to carriers, once the application is filed with the carrier, the UPIN assignment process should take two weeks or less, on average.

1.2.2 Use of UPINs on Medicare Claims

The HCFA-1500, or the electronic media claims format, should be used by all physicians and entities billing Medicare. The health insurance claim form includes fields for UPIN information for the referring physician and the carrier assigned PIN for the performing physician.

1.2.2.1 Performing Physician UPIN

Items 31 and 33 on the HCFA-1500 identify the physician/supplier who renders a service or item to a Medicare beneficiary. The physician (or representative) is required to record the billing name, address, zip code, phone number, and PIN number (or group number) in Item 33. Item 31 requires the signature of the physician/supplier or representative including degrees or credentials. There is no field for the performing physician UPIN on the claim form. Individual rendering physicians are asked to provide their PIN number in Item 24K, next to each line item service rendered. Group members should have PINs identifying them within the group, so that each rendering physician is uniquely identified, regardless of group member status. Carriers are instructed to cross-reference carrier issued identification

numbers (PINs) with HCFA-issued UPINs for performing physician services and to include the UPINs on Part B claim records before transmitting records to the CWF (Program Memorandum B-91-12).

1.2.2.2 Referring Physician UPIN

Effective January 1, 1992, a physician or supplier who bills Medicare for a service or item must show the name and the UPIN of the ordering/referring physician on the claim form, if that service or item was the result of an order or referral from a physician (Program Memorandum B-91-12). On the HCFA-1500, the name and the UPIN of the referring physician must be entered in Items 17 and 17a, respectively for selected services. On electronic claims, the name and UPIN must be entered in Record/field EAO-20.0, positions 80-94.

To facilitate the reporting of these UPINs, OBRA90 required HCFA to publish a UPIN Directory of all physicians registered with a carrier and with the Registry. Carriers were to distribute these directories (organized by carrier number) upon request and free of charge to physicians, DME suppliers, laboratories, hospitals, ESRD facilities, radiology and other imaging centers, physical therapy facilities, consulting physicians, and billing services submitting claims for these entities. A March 1992 Program Memorandum also instructed carriers to notify physicians through state and local medical societies and trade associations that they must obtain a UPIN if they refer Medicare patients, even if they never bill Medicare directly.

Carriers were instructed through Program Memoranda (December 1991 and March 1992) in HCFA policy for editing/rejecting claims without the referring/ordering UPIN. As of January 1, 1992, carriers were instructed to reject claims that do not contain the name and the UPIN of the referring/ordering physician (or an acceptable surrogate as discussed below) for:

- consultative services (TOS 3)[4];
- diagnostic radiology services (TOS 4);
- diagnostic laboratory services (TOS 5);
- specific DME (including orthotics and prosthetics).

All other services should also include the UPIN of the referring/ordering physician when a referral is involved; however, claims for other services are not being edited by CWF at this writing. If a claim was self-referred, referred by a non-physician (e.g., pharmacist, midwife), the performing physician should report his or her own UPIN in the spaces provided. Also, in the case of a second opinion, the consulting physician would show his or her own UPIN instead of the provider of the first opinion because a second opinion is initiated by the beneficiary.

Only the initial claim of the result of a referral is required to show the referring/ordering physician. In other words, after the initial referral, a consulting physician would become the ordering/referring physician as well as the performing physician on all subsequent visits. In the case of a referral to a laboratory or DME supplier, subsequent visits to the laboratory should include a referring/ordering physician UPIN because these services require referrals.

[4]Previously, the services to be edited for the referring/ordering physician were defined by specialty codes. However, because of the "potential conflict between carrier edits based on specialty codes, and CWF edits based on procedure codes" the March 1992 Program Memorandum defined the services requiring edits by CWF type of service (TOS) codes.

1.2.2.3 Surrogate UPINs

Surrogate UPINs were designed to be used by the referring physician and only if that physician has not yet been issued a UPIN. Surrogate UPINs contain alpha characters in the first three positions and numerics in the last three positions. They are available for:

- interns and residents (INT000 or RES000);
- retired physicians referring services (RET000);
- physicians with the Military or Veteran Affairs (VAD000);
- physicians with the Public Health Service (PHS000); and
- physicians who do not have a UPIN and who does not qualify for one of the above categories (OTH000).

All of the above surrogate UPINs are temporary until a physician receives a UPIN, except for RET000. However, a retired physician furnishing a Medicare service must obtain a UPIN. The March 1992 Program Memorandum instructed carriers to discontinue the use of the surrogate SLF000 because for a self-referred or non-referred service or visit, the referring UPIN should be that of the performing physician. The OTH000 surrogate UPIN should only be used when a referring/ordering physician has not been assigned a UPIN and does not qualify for one of the other surrogates.

1.2.2.4 Claim Edit Process for UPINs

Carriers were required as of January 1, 1992, to reject any claim for consultative services, laboratory services, diagnostic radiology services, and DME, that does not include the UPIN of the referring/ordering physician. However, after August 1, 1992, CWF became

responsible for all of the editing (March 1992 Program Memorandum). Claims that do not meet the referring UPIN requirement will be returned to the carrier.

Carriers are instructed to return, deny, reject, or develop a claim that is returned by CWF, depending on their system's capability. If developing is more cost effective for a carrier it may be done, but according to the March 1992 Program Memorandum, HCFA feels that developing is "least effective in terms of obtaining compliance with the requirements." If a carrier denies a claim, it must: 1) explain why the claim was denied, 2) ask that the physician resubmit the claim with the missing UPIN information, and 3) offer an appeal. No CWF edits for the performing physician UPIN exist at this writing.

2.0 REVIEW OF SELECTED REGISTRY AND CARRIER OPERATIONS ISSUES

This chapter summarizes selected issues in the operations of the Physician Registry contractor, Transamerica Occidental Life Insurance Company (TOLIC), and in the UPIN operations of carriers. The discussion focuses on operations aspects as they bear on the quality of Registry data. Chapter 1 described the basic relationships between the Registry and the carriers. In some respects, Chapter 1, based primarily on official documents, has presented a theoretical view of how the UPIN database should be developed and maintained. This chapter, in contrast, describes actual problems experienced during the implementation phase, and on-going problems and procedures. The first section of this chapter reviews TOLIC's Registry operations. The second section reviews carrier operations, except for the two Railroad Retirement Board (RRB) carriers. The last section describes RRB carrier operations.

This chapter is almost entirely based on site visits conducted during August and September, 1993 – a visit to the TOLIC Physician Registry in Los Angeles and six site visits to Medicare carriers. The carriers visited were (1) C & S Administrative Services, which covers Maine, Massachusetts, New Hampshire, and Vermont; (2) TOLIC in its capacity as a carrier for southern California; (3) Pennsylvania Blue Shield, which covers Delaware, the District of Columbia, New Jersey, and Pennsylvania; (4) Illinois BC/BS; (5) Florida BC/BS; and (6) the Travelers Insurance Company's RRB operations in Georgia. Additional information came from follow-up phone calls.

Before describing the individual participants' experience during the implementation phase of the Registry and current operations, an overview of the implementation phase of the Registry is provided.

2.1 Implementation Phase of the Physician Registry[1]

The implementation phase of the Physician Registry occurred during the five month period from January through May 1989. TOLIC had to process over 1.2 million records submitted by 57 carriers[2] during this period. Several of the required data elements for the Physician Registry had not been collected by some carriers prior to the establishment of the Registry. Carriers, however, were instructed by HCFA not to directly contact providers to obtain values for required data elements that the carriers did not have in their provider files.[3] Nor were carriers allowed to obtain such information from copyrighted sources. For the date of birth, a required data element, carriers were allowed to submit just the year portion of the date if they did not have the complete date. Other data elements (i.e., medical school code, graduation year, and state license number) were not mandatory during the implementation phase. Since carriers were not provided additional funding to develop incomplete records, it was understood that carriers would be submitting records with incomplete information to TOLIC.

2.2 Registry Contractor

TOLIC is responsible for incorporating the provider data submitted by carriers into the Registry. Before incorporating new data into the Registry, TOLIC screens the data elements using what it calls "online edits" (Exhibit 2-1). The *on-line edits* performed by TOLIC have evolved over time. However, it appears from the data in Chapter 4 that errors exist in the Registry data even for practice settings added as recently as 1992 and 1993. For instance, 1,025

[1]Some of the material in this section is drawn from a letter from Transamerica to HCFA (March 4, 1994).

[2]The two RRB carriers did not submit records to TOLIC during the initial phase.

[3]This restriction was removed after the implementation phase.

EXHIBIT 2-1

REGISTRY DATA EDITS (October 19, 1993)

UPIN ON-LINE EDITS

FIELD	DATA NAME	EDITS
001	CAR-RECORD-CODE	Must be 1, 2, 5 or 7.
002	CAR-PHYSICIAN-STATUS	Must be 1 when CAR CREDENTIALS = MD Must be 2 when CAR CREDENTIALS not = MD
003	CAR-LAST-NAME	Must not be BLANK.
004	CAR-FIRST-NAME	Must not be BLANK.
008	CAR-CITY1	Must not be BLANK.
009	CAR-STATE1	Must be a valid state.
010	CAR-ZIP-AND-SFX1	Must be greater than ZERO.
013	CAR-STATE2	Must be a valid state.
015	CAR-STATE-LICENSED	Must be a vaild state.
016	CAR-STATE-LIC-NUMBER	Must not be ZERO or BLANK.
017	CAR-DATE-OF-BIRTH	Must be numeric, month cannot be greater than 12, day cannot be greater than 31.
018	CAR-MED-SCHOOL-CODE	Must be a valid school code and must be valid for the credential in CAR-CREDENTIALS. MD must have an MD school, DPM must have a DPM school, etc.
020	CAR-DATE-OF-DEATH	If not ZERO, month must be less than 13 and day must be less than 32.
021	CAR-CREDENTIALS	Must be a valid credential, must be MD or DO when CAR-PHYSICIAN-STATUS = 1 and other than MD or DO when CAR-PHYSICIAN-STATUS = 2.
022	CAR-PRIMARY-SPECIALTY	Must be a valid specialty.
023	CAR-PRIMARY-BD-CERT	Must be a valid certification code.
024	CAR-SECOND-SPECIALTY	Must be a valid specialty.
025	CAR-SECOND-BD-SPECIALTY	Must be a valid certification code.
026	CAR-SANCTION-CODE	Must be a valid sanction code when not BLANK.
027	CAR-EFFECTIVE-DATE	Must be numeric when not ZERO.
028	CAR-NO-YRS-SANCTION	Must be numeric when not ZERO.
029	CAR-RES-INTERN-CODE	Must be a valid indicator. Must be equal to D when CAR-DATE-OF-DEATH is greater than ZERO.
030	CAR-GROUP-PRACTICE-IND	Must be a valid group indicator.
031	CAR-PARTICIPATING-IND	Must be a valid indicator.
033	CAR-PROVIDER-NUMBER	Must not be BLANK or ZERO.
034	CAR-UPIN-ERROR-CODE	Must not be BLANK or ZERO when CAR-MPIER-UPD-FLAG = Y. Must be given for CAR-RECORD-CODE = 5 or 6.
035	CAR-CONTROL-NUMBER	Must not be ZERO or BLANK when CAR-RECORD-CODE = 2 or 6.
036	CAR-CARRIER-NUMBER	Must be a valid carrier.

EXHIBIT 2-1 (continued)

REGISTRY DATA EDITS (October 19, 1993)

UPIN ON-LINE EDITS

FIELD	DATA NAME	EDITS
039	CAR-MPIER-UPD-FLAG	Must be a valid flag. CAR-RECORD-CODE = 1 or 2 may have a value or BLANK or Y. If Y CAR-UPIN-ERROR-CODE must be present. CAR-RECORD-CODE = 5 or 6 may have a value of BLANK, 1 or 2. If 1, CAR-SPEC-PROCESS-DATA must contain physician's last name as it appears on the MPIER file. If 2, CAR-SPEC-PROCESS-DATA must contain the physician's old provider number as it appears on the MPIER file. CAR-UPIN-ERROR-CODE must be present.
040	CAR-SPEC-PROCESS-DATA	Must not be BLANK when CAR-RECORD-CODE is equal to 5 or 6 and CAR-MPIER-UPD-FLAG is equal to 1 or 2.

EDITION: 10/19/93.

Source: Transamerica Occidental Life Insurance Co.

erroneous values of birth year were found in practice settings added to the Registry during 1992-93.

For any given provider, TOLIC maintains one record in the Registry. The record consists of a *header* section, containing data that should be invariant across practice settings, and a *practice setting* section. Within the practice setting section, there are as many practice setting segments as the provider has practice settings, both active and inactive. Data in the header section are derived from the practice setting information on a flow basis as submitted by carriers on add or update records. The first practice setting entered into the Registry was the source of the original header data.

As used in actual operations, a "practice setting" is usually a specific location at which a physician renders service, and can be identified as physically separate from any other service location. In addition, when an individual physician practices under more than one organizational arrangement, e.g., group practice and solo, there will be a practice setting record for each. This would be true even if the group and solo physical locations are one and the same. A separate practice setting may also be defined for each payment locality in which a physician practices, even if all activities are conducted at a single site. (A payment locality is a geographic or other unit created for pricing purposes.) The rules for updating header information follow.

Several data elements are common to both the header and practice setting segments of the provider record (see Exhibit 3-1). Of the common data elements, only the name fields are not automatically updated on the header segment when new feasible values are available on the incoming carrier record. For the other common data elements (date of birth, date of death, medical school code, graduation year, and the three sanction data elements), whenever a new or updated practice setting is submitted with an UPIN to the Registry, any new feasible

(practice setting) value automatically replaces an old header value, regardless of the feasibility of the old value.[4]

For carrier records submitted without a UPIN, TOLIC will try a computerized match against existing header information. The match is based on a key constructed out of the first three letters of the provider's last name, the first letter of the first name, a "soundex" value based on the remaining letters of the last name and, in some cases, a zip code. If the "keys" match, then the computer algorithm checks practice setting information: the provider's last name, first name, middle initial, full date of birth, school code, graduation year, and state licensed in all must exactly match in order to attach the practice setting to an existing provider record. When checking the provider's existing practice settings, the algorithm checks one setting at a time. It does not borrow information that matched from previous settings that were checked. If a match is not found, TOLIC then checks an American Medical Association (AMA) data extract. If a match still has not been found, TOLIC's system will flag the new practice setting for manual review. The purpose of the review is to determine whether to attach the incoming information to an existing provider (UPIN) or whether a new UPIN needs to be provided. A TOLIC analyst will review existing UPIN records with the provider's last name providing the basis for the initial screening element. When reviewing records, the analyst has full access to both header and practice setting information.

If before updating the older header value was feasible, it originated from one of the previously submitted practice settings; thus the new header value must conflict with at least one of the feasible practice setting values. TOLIC, *per se*, does not resolve differences in practice setting values. One reason TOLIC does not resolve differences is due to the belief that more recent values are more accurate than older values. Also, the means by which TOLIC might resolve discrepancies are limited. For instance, if two or more carriers have submitted

[4]The provider's last name and the first letter of the first name need to match before the other elements are automatically updated.

inconsistent information and all of them claim that their information is correct, TOLIC usually does not have the means to decide which values are correct.

TOLIC reported that some carriers seem to believe that it is the responsibility of the Registry to detect errors in their submittals. Our interviews with the case study carriers suggest that this perception is accurate. Such confusion about the respective roles of TOLIC and the carriers in ensuring data integrity may lead to weaknesses in the Registry data base.

2.3 Carriers

All of the interviewed carriers indicated that current operations were running smoothly. Many problems, however, were encountered during the first few years of operation. Indeed, many of the data problems detailed in Chapter 4 originated in the implementation phase. Several Registry data elements required by HCFA had not been previously collected by the carriers. For the initial batch submission of data to the Registry, therefore, the carriers were directed to obtain new, required data elements. Because of HCFA regulations, however, the carriers were not allowed during the implementation phase to directly contact providers for the requisite information.[5] Carriers also were not allowed to use copyrighted American Medical Association (AMA) data. These restrictions frequently led to missing or erroneous data, because carriers lacked resources for tracking down information in other ways. Other weaknesses in Registry data stem from data elements which HCFA initially stipulated as optional, but which became mandatory in December 1991, namely, state license number, medical school, and graduation date.

Some carriers faced additional start-up challenges. For instance, just as the implementation was commencing, Illinois BC/BS took over from EDS the carrier contract for Illinois. Simultaneously, Illinois BC/BS adopted an entirely new claims processing system, one

[5]The regulations were modified at a later date to allow carriers to directly contact providers.

of eight shared systems. During the late 1980s and early 1990s, HCFA required that carriers adopt a shared system (Exhibit 2-2).[6] In most instances, this meant that carriers utilize a standard package of software to maintain their provider files and process claims. For Illinois BC/BS, this meant that they had to convert EDS provider identification numbers (PINs) to their own system and then reconvert them to the new shared system format. Pennsylvania BS likewise had problems with their operations when they assumed the New Jersey contract after UPIN implementation had commenced. Pennsylvania BS had to rectify deficiences in the New Jersey data left by the former carrier. Because we only conducted a few site visits, we do not know to what extent other carriers experienced similar problems (e.g., a simultaneous assumption of a carrier contract or simultaneous entry into a shared system).[7]

Each carrier maintains a provider file that is the source of data submitted to the Registry. When adding a new practice setting for a provider that is already present in the provider file, information from the *header* section, if such a section exists, is automatically filled for the new practice setting. Data in the *header* section consists of data elements whose values should not vary by practice setting (e.g., date of birth). If there is no valid value present for a required data element, either in the existing header section or on the application form for a new practice setting, carriers are allowed to directly contact the provider for the missing information.

The extent to which carriers screen or edit data before sending it to the Registry varies. Most carriers apply several edits and screens, such as for valid date formats. Some carriers, however, have only minimal on-line edits. In some instances, however, the *on-line* edits are quite extensive and comprehensive (e.g., Florida BC/BS). Florida has installed edits that will not allow an invalid medical school code.[8] To identify data problems not caught during entry

[6]Pennsylvania BS is its own shared system.

[7]According to Exhibit 2-2, ten carriers entered a shared system during 1988 and 1989 and another nine during 1990. However, it is not clear to what degree the planning and implementation of the Registry and the shared systems interfered with each other.

[8]In some instances, the on-line edits of Florida's system are too restrictive for Registry purposes and Florida edits a down-loaded version of the file containing new information before sending it to the Registry.

EXHIBIT 2-2

CURRENT SHARED SYSTEMS ARRANGEMENTS (September 28, 1993)

PART B SYSTEMS

SYSTEM NAME	OPERATIONAL SITES	DATE
Optimum Systems Inc. (OSI/Shared Arkansas System (SAS))	Arkansas (00520)/Louisiana (00528)	01/85
	Minnesota (00720)	05/90
Proprietary (Maintained by Arkansas Blue Shield)		
Multi-Carrier System	Massachusetts/Tri-State-SP	07/88
	Mass B/S (00700)	
	New Hampshire/Vermont (00780)	
Proprietary (Maintained by Electronic Data Systems-Federal (EDSF))	Maine (21200)	
	HCSC-SP Health Care Services Corp. [aka Illinois B/S (00621)]	04/89
	Western New York-SP (00801)	11/89
	Nationwide-SP	12/89
	Ohio (16360)	
	West Virginia (16510)	
	California-SP (Cal B/S-00542)	08/90
	Texas (00900)	12/92
GTE Medicare System (GTEMS)	Florida (00590)	12/88
	Utah (00910)	10/90
	Wisconsin Phys. Serv. (00951)	03/91
Proprietary (Maintained by GTE Data Services)	Aetna	08/92
	Alaska (01020)	
	Arizona (01030)	
	Georgia (01040)	
	Hawaii (01120)	
	Nevada (01290)	
	New Mexico (01360)	
	Oklahoma (01370)	
	Oregon (01380)	
	**Washington (01390) 1/1/94	N/A

EXHIBIT 2-2 (continued)

CURRENT SHARED SYSTEMS ARRANGEMENTS (September 28, 1993)

PART B SYSTEMS

SYSTEM NAME	OPERATIONAL SITES	DATE
Metropolitan Medicare System Proprietary (Maintained by Viable Information Processing Systems (VIPS))	Kentucky (00660) Puerto Rico & VI (00973) Indiana (00630) CGLIC (Conn Gen Life) Idaho (05130) Tennessee (05440) North Carolina (05535) Group Health Ins. (NY 14330) South Carolina (00880) Michigan (00710) Alabama (00510) Kansas-SP (00650) (Nebraska 00655) (Kansas City 00740) Empire NY (00803) Transamerica Cal (02050)	10/82 03/85 06/88 12/88 10/89 03/90 08/90 11/90 10/91 10/91 11/91 06/92
HCFA Part B Standard System (HPBSS) Public Domain (Maintained by the Travelers)	Travelers-SP Connecticut (10230) Minnesota (10240) Mississippi (10250) Virginia (10490) (Railroad Ret. Board 10072) (Rhode Island 00870)	06/88 06/88 09/91
21st Century Medicare System-B (21MSB) Proprietary (Maintained by North Dakota Blue Shield)	North Dakota/S. Dakota-SP (00820) (Colorado 00550) Iowa (00640) Wyoming (00825)	05/90 10/92 07/91

EXHIBIT 2-2 (continued)

CURRENT SHARED SYSTEMS ARRANGEMENTS (September 28, 1993)

PART B SYSTEMS		
SYSTEM NAME	OPERATIONAL SITES	DATE
BMACS (Part B Medicare Automated Claims System)	King County-SP Washington (00932)	08/91
	(Montana 00751)	08/91
Proprietary (Maintained by King County** Medical Blue Shield)		
GAMSS (General American Medicare Standard System)	Gen American Missouri (11260)	02/90
	Maryland (00690)	10/91
Public Domain (Maintained by General American)		
EXEMPTED SYSTEM	OPERATIONAL SITE	DATE
	Pennsylvania BS Pennsylvania (00865) Delaware (00570) DC (00580) New Jersey (00860)	N/A

Notes:

N/A = Not applicable.

SP = Contractor responsible for shared processing of contractor in parenthesis below it; with EDS contractors, none except TX has a copy of the EDS software on site.

**King County will lose its carrier contract with HCFA as of January 1, 1994. Washington's new carrier will be Aetna; the new carrier number will be 01390. Montana (00751) will migrate to GTE shared system as a stand alone carrier.

Source: Health Care Financing Administration.

into their provider file, some carriers (e.g., Pennsylvania BS) perform post-entry data edits before sending the data on to the Registry. At least one carrier, Pennsylvania BS, performs routine computer runs to detect the existence of problems in their Medicare provider database. The results are then used to create update transmission records for the Registry.

The operations by which the carriers process add records, changes to existing practice settings, and other changes are evolving as carriers gain more experience in maintaining their provider files. For carriers on shared systems, however, the other members of the shared system must agree to any changes. If a carrier wants a change that the other carriers do not want, the requesting carrier has to incur all costs for changing its version of the software. Carriers are not supposed to have customized versions of the software made for them.

All of the carriers indicated that it would be an enormous and costly effort to retrospectively correct data problems existing in their current provider files, even for active practice settings only. The single biggest problem in trying to correct existing data is that carriers have leverage on providers to submit the required data *only when* the provider initiates a new practice setting application.

2.4 Railroad Retirement Board (Georgia)

The RRB carrier operations are unique within the Medicare system and, as such, present special challenges. In particular, RRB carriers do not have the same access as other carriers to high quality provider information. As a consequence, RRB carriers are more likely than other carriers to submit inaccurate and/or incomplete carrier records to TOLIC. Because of the large number of providers they service, the RRB carriers are often associated with a disproportionately large share of the missing values and erroneous values found in Registry data. For instance, the overall missing/erroneous value rate for birth year in 1992-93 for all carriers was 5.2 percent. For the two RRB carriers the rate was 15.2 percent while for the other

carriers, the combined rate was 0.33 percent. The rest of this section describes the reasons the RRB carriers have difficulties obtaining quality information. In addition, RRB attempts to change their relationship with the Registry are described.

As a system, the RRB predates Medicare. RRB beneficiaries, unlike the beneficiaries of other carriers, are drawn from throughout the country, and so are their providers. Whenever a RRB beneficiary obtains medical care, the claim is submitted to the RRB rather than to the usual Medicare carrier for their state. This situation presents the RRB carrier with added difficulties in obtaining and validating UPIN information. Not only is their UPIN workload extraordinary in size, but at the same time they are farther than other carriers from the necessary sources of information.

The RRB often has difficulties obtaining data from providers. The main reason is that most providers do not provide care for many RRB beneficiaries. For this reason alone, providers do not have much incentive to fully cooperate with the RRB. RRB relations with providers are aggravated by Medicare reimbursement rules. In contrast to normal situations, the reimbursement amount is determined by beneficiary's residence rather than the provider's location. Further, if participating, then the provider must accept assignment.

In addition to difficulty in obtaining information from providers, data entry errors can arise because providers do not need submit RRB application forms. That is, the RRB can not insist that providers use RRB application forms for new practice settings, because HCFA allows providers, instead, to submit copies of applications previously submitted to their normal carrier(s). Because application forms vary by carrier, this inevitably gives rise to situations in which data entry errors can easily occur. Another problem, at least in the past, was that Medicare participation status collected by other carriers had to be manually entered by the RRB carriers into their provider files. Only recently were the RRB carriers able to receive data tapes directly from the other carriers.

Aside from the difficulties in obtaining information confronting the RRB carriers, the participation of the RRB in the Registry leads to a duplication of effort. In theory, the RRB carriers could have in their provider files every active practice setting in the nation. At a maximum, then, two-thirds of all active practice settings in the Registry would be from the two RRB carriers. Although the theoretical maximum has not been approached, the two RRB carriers are ranked first and third among carriers in terms of the number of practice settings maintained on the Registry. This amounts to a large duplication of effort.

The RRB carriers recognized that their participation in the Registry would be a duplication of effort. As a consequence, they made two attempts to change their relationship with the Registry. The RRB requested that the Registry send a tape containing Registry data to them, in order to maintain RRB records. (RRB then would not need to otherwise participate in Registry activities.) Failing that, the RRB would have liked to obtain, via on-line data searches of the Registry, the values of data elements missing in their application forms. HCFA turned down both of the RRB's requests. At present, because of HCFA regulations regarding timeliness of application processing, the RRB forwards incomplete records to the Registry even though it knows that the Registry will flag the data because of erroneous values or other data problems.

3.0 ANALYSIS OF PHYSICIAN REGISTRY DATA

In this chapter we describe the characteristics of the file used in the analyses, the classification scheme developed to characterize data element values, and the cross-classification (control) variables used for analyzing the data.

3.1 File Description

The data used in the analysis were sent to HCFA by the contractor, Transamerica-Occidental Life Insurance Company, Inc. (TOLIC), that maintains the Registry. TOLIC sends an updated version of the Registry data to HCFA on a quarterly basis. The July 1993 version was used in our analysis.

As received by HCFA, the file consists of variable-length records. There is one record per provider on the file. Each record has a *header* section followed by a *practice setting* section that contains data on one or more practice settings for that provider. Since the number of practice settings varies by provider, the length (size in bytes) of the *practice setting* section of the record varies in direct proportion to the number of practice settings.

For our analysis, we used a HCFA version of the Registry data in which the variable-length records had been converted to fixed-length records. Each fixed-length record consists of two major sections: (1) data from the *header* section of TOLIC's file and (2) data from *one* of the practice settings. In the file with fixed-length records, there are as many records for a given provider as the provider has practice settings.

There is some overlap in the data elements contained in the header and practice settings sections of each record. Provider name information, sanction information, dates of birth and death, medical school, and year of graduation are present in both sections. As noted in Chapter 2, the values of common data elements may differ in the two locations (i.e., header

versus setting). In addition to the common data elements, Exhibit 3-1 indicates the location of other data elements in the provider record.

3.2 Classification of Data Element Values

The purpose of this report is to investigate the integrity of Registry data. To facilitate the investigation, a classification system for the values of data elements was developed. The system classifies a data element's value into one of five categories; it is briefly described below. Following the description of the classification system, for each data element analyzed in Chapter 4, the criteria for determining how values were classified into one of the five categories are described.

3.2.1 Classification System for Data Element Values

The values for each data element were generally classified into one of five categories: (1) feasible, (2) missing, (3) erroneous, (4) dubious, and (5) unknown. In general, HCFA's specifications were followed in classifying values into one of the five categories. There were instances, however, in which other values (sometimes nonsensical) appeared in the data. In some cases, such values were not automatically classified as erroneous but rather were reviewed to determine whether another category would be appropriate.[1] For example, the appearance of blanks in numeric fields[2] was usually classified as missing rather than erroneous. When a data element value did not meet the criteria for one of the other four categories, it was automatically classified as erroneous.

[1] Also see the discussion on the "year setting added" data element.

[2] A numeric field should ordinarily be filled with zeroes if the data point is missing or the field is otherwise intended to be null.

EXHIBIT 3-1

REGISTRY DATA ELEMENTS

Data Element	SECTION OF EACH RECORD	
	Header	Setting
Physician last name	X	X
Physician first name	X	X
Physician middle name	X	X
Physician suffix	X	X
Previous physician last name	X	X
Previous physician first name	X	
Previous physician middle name	X	
Previous physician suffix	X	
Date of birth	X	X
Date of death	X	X
Medical school	X	X
Year graduated	X	X
Sanction code	X	X
Sanction, effective date	X	X
Sanction, number of years	X	X
State license number		X
Physician status		X
State licensed in		X
Credentials		X
Primary specialty		X
Primary board certification		X
Secondary specialty		X
Secondary board certification		X
Date setting was added		X
DRIP status		X
Group practice indicator		X
Participation indicator		X

Note: X indicates location of data elements.

For data elements which are coded (e.g., sanction code and medical school code), only those values which are on HCFA's official list of codes are considered *feasible*. For date-type variables, criteria for determining feasible values were developed for each variable. Note that a feasible value is not necessarily an accurate value. For example, a date of birth equal to February 2, 1953, is valid, in the sense of being a feasible value. But for a given provider, a "feasible" date might be incorrect. We were not able to check the data for this type of accuracy. Because data formats in the file that HCFA receives from TOLIC sometimes differ from the records that carriers send to TOLIC (see Exhibit 3-2), we sometimes had to employ several extra steps of logic to differentiate among the five categories.

When values were unavailable (missing) for a data element, HCFA instructed carriers to fill numeric fields with zeroes and alphanumeric fields with blanks. In most instances, our definition of *missing* follows HCFA's instructions. In some instances, if a numeric field was filled with blanks instead of zeroes, it is also considered missing. Other exceptions are described in the next section.

For code-type variables, *erroneous* refers to invalid codes or codes which are not applicable to UPIN providers. For other variables, it refers to instances in which the values are not credible. For instance, since our file was current through mid-1993, the year portion of the date of death could not have values greater than 93.

There are some values of data elements which are simultaneously feasible and erroneous; such values are considered *dubious*. A date of death in 1982 is an example. The date, in itself, is feasible. Carriers, however, were instructed only to submit practice settings that had claims activity during the two years prior to the Registry start-up. Many carriers sent their initial set of practice settings to the Registry in early 1989. This means that only practice settings with claims activity in 1987 and 1988 should have been sent to the Registry. Since it is quite unlikely that there were outstanding claims from 1982 still to be adjudicated in 1987, a date of death of 1982 is in a sense erroneous because the practice settings of providers they

EXHIBIT 3-2

FORMATS OF REGISTRY DATA ELEMENTS

	Carrier Record					*Registry File*	
Data Element	Required	COBOL Picture	Justifi-cation	Component Order	Missing Value	COBOL Picture	Component Order
Date of birth	Y	9(06)	R	MMDDYY	0	S9(09) COMP-3	YYMMDD
School code	Y	X(05)	R		blank	X(05)	
Year graduated	Y	9(02)	R	YY	0	S9(09) COMP-3	YY
Sanction code	N	X(01)			blank	X(01)	
Sanction, effective date	N	9(04)	R	MMYY	0	9(04)	MMYY
Sanction, number of years	N	9(02)	R		0	S9(03) COMP-3	
State license number	Y	X(12)	R		blank	X(12)	
Physician status	Y	9			0	X(01)	
Provider credentials	Y	X(03)	L		blank	X(03)	
Primary specialty	Y	9(02)	R		0	X(02)	
Secondary specialty	N	9(02)	R		0	X(02)	

Sources:

1. Carrier record formats from HCFA Transmittals.
2. Registry file formats from Transamerica Occidental Life Insurance Company: version of file sent to HCFA.

represent should not have ever been included in the Registry. Rather than classify 1982 as feasible or erroneous, the *dubious* category was devised to handle such cases. *Dubious* is also applied to situations in which a value is feasible but is of marginal credibility.

The final category is the *unknown* category. This category is only used for four of the code-type data elements: the two sanction code data elements and the two specialty code data elements. For these data elements, a code for "unknown" is in HCFA's official list of codes for the purpose of indicating that the correct value is not known to the carrier. Although valid codes, they indicate indeterminacy in the data and, hence, are summarized in the tables.

3.2.2 Classification of Individual Data Elements

From the date of birth, only birth year is examined in Chapter 4 because HCFA allows carriers to submit just the year portion of the date. To be considered feasible, a birth year had to be between 1900 and 1969, inclusive. A birth year of 1899 is considered dubious as long as it has a feasible month value (i.e., 1-12). If the entire date of birth field is zero-filled, blank-filled, or filled with nines (9), it is considered missing.[3] All other year values are considered erroneous. That is, dates of birth after 1969 are not considered feasible. In other words, for the birth year to be feasible in 1993, a provider has to be older than 24 years old and less than 94 years old. Except for 1899, no other year was considered dubious. Since it is unlikely that many providers 80 years old or older (in 1993) are still practicing, a case can be made that dates of birth prior to 1914 should also be considered as dubious. However, it is known that there are some medical doctors 80 years old or older still practicing. It was finally decided to only include 1899 in the dubious category. The classification of the values of birth year and other

[3]Although date of birth is a numeric field, we considered blanks as indicative of missing values. A couple of carriers apparently mis-interpreted HCFA's instructions and filled nines instead of zeroes to indicate missing values. We considered a nine-filled field as missing.

data elements are summarized in Exhibit 3-3. (Because year is represented in the data in a YY format rather than a YYYY format, only the significant YY values are shown in Exhibit 3-3.)

From the date of death, only death year is examined in Chapter 4 because HCFA allows carriers to submit just the year portion of the date. Since our file was current through mid-1993, the year portion of the date of death could not have values greater than 93. The 1986 year of death cutoff was determined through a more complex process. First, we took into consideration that carriers were instructed during the implementation phase (late 1988 and early 1989) of the Registry to only submit data on practice settings in which the setting had claims activity during the two previous years (i.e., 1987-8). However, because some claims continue to be adjudicated after a provider's death, claims activity for a provider who died in 1986 might have continued into 1987. Thus, 1986 was allowed to be an acceptable year of death. Hence, the feasible range of death year is 1986 to 1993, inclusive.

Dates of death prior to 1986 and after 1900 were considered dubious. In the case of dates of death prior to 1986, it could easily be the case that the date is accurate. However, such information is not supposed to be in the Registry, given the criteria by which carriers were to submit data. If the entire date of death field is blank-filled it is considered missing. If the date of death field was filled with zeroes or nines, for this report, the provider was considered as still alive. Given HCFA's instructions for date of death, a live provider and a dead provider for whom a date of death is unavailable can not be distinguished.[4]

Only codes on the official list are considered feasible for medical school code. If the school code is equal to all zeroes, "00001," or blank, then the school code is considered missing. All other values are considered erroneous.

For graduation year, feasible years are from 1940 to 1993, inclusive. If the field was filled with zeroes or nines, then it is considered missing. The graduation years 1930 to 1939 are

[4]Carriers were instructed to indicate a death by replacing a zero-filled field with the date of death. HCFA's coding instructions made no provision for a situation in which a provider is known to have died but the date of death is unknown.

EXHIBIT 3-3
DEFINITION OF DATA ELEMENT INTEGRITY MEASURES

Data Element	Category	Criteria*
Birth Year (Header & Setting)	Feasible	(1≤Year≤69) or (Year=0 and (1≤Month≤12) and (day value consistent with month))
	Dubious	Year=99 and Month≤12
	Missing	All zeroes or all nines or blank
	Erroneous	All other values
Death Year (Header & Setting)	Feasible	(86≤Year≤93) or all zeroes or all nines
	Dubious	(1≤Year≤85)
	Missing	Blank
	Erroneous	All other values
Medical School (Header & Setting)	Feasible	Only codes on Registry list
	Missing	Blank or all zeroes or 00001
	Erroneous	All other values
Graduation Year (Header & Setting)	Feasible	40≤Year≤93
	Dubious	30≤Year≤39
	Missing	All zeroes or all nines
	Erroneous	All other values
Year Sanctioned (Header & Setting)	Feasible	((66≤Year≤93) and (Month≤12)) or all zeroes
	Missing	Blank
	Erroneous	All other values
Sanction Code (Header & Setting)	Feasible	Codes A-R or blank
	Unknown	Code U
	Erroneous	All other values
Sanction Length (Header & Setting)	Feasible	0, 1-15, 20, 25, 30, 35, 40, 45, 50, or 99
	Dubious	16-19, 21-24, 26-29, 31-34, 36-39, 41-44, 46-49, or 51-98
	Missing	Blank
	Erroneous	All other values
Primary Specialty (Setting)	Feasible	01-41, 44, 46, 48, 66, 70, 76-79, 81-86, or 90-94
	Missing	All zeroes or blank
	Unknown	99
	Erroneous	All other values
Secondary Specialty (Setting)	Feasible	0, 01-41, 44, 46, 48, 66, 70, 76-79, 81-86, or 90-94
	Missing	Blank
	Unknown	99
	Erroneous	All other values
State Licensed In (Setting)	Feasible	Standard state postal code
	Missing	Blank
	Erroneous	All other values
State License Number (Setting)	Missing	Blank
	Feasible	All other values
Physician Status (Setting)	Feasible	1 or 2
	Erroneous	All other values
Primary and Secondary Board Certification (Setting)	Feasible	Blank, Y, N, or U
	Erroneous	All other values
Group Practice Indicator (Setting)	Feasible	1 or 4
	Missing	Blank
	Erroneous	All other values
Participation Indicator (Setting)	Feasible	Y or N
	Erroneous	All other values

*Year = year portion of date; month = month portion of date; Y = yes; N = no; U = unknown.

considered dubious. A person graduating from medical school in 1930 at the age of 22 would, in 1993, be 85 years old and, in 1988, would have been 80. The probability that physicians between ages 76 and 85 are still practicing medicine is low; therefore, we took a conservative approach, judging these cases warranting caution on the part of potential users of the data. Note that retired providers may use a special surrogate UPIN for use in claims processing (if they have not obtained a UPIN) and that by the 1930s, the typical age at graduation from medical school was more likely to have been 25 or 26 than 22.[5] All other graduation year values are considered erroneous.[6]

For date sanctioned, a data element that included month as well as year, the year portion had to be equal to 66 to 93 (inclusive) and the month portion had to be equal to 12 or less in order to be considered feasible.[7] Dates prior to the establishment of Medicare (1966) were not considered feasible. Like date of death, a zero-filled date sanctioned field indicates a non-sanctioned provider or a sanctioned provider whose date of sanction was not known by the carrier. In this case, a zero-filled field is considered feasible. A blank-filled field is considered missing. All other values are considered erroneous.

A code had to be on HCFA's official list (A-R) or be blank to be considered feasible for sanction code. A blank indicates that the provider is either not sanctioned or is sanctioned but the carrier does not know which code to assign (despite having U available to indicate "unknown"). If the sanction code is equal to U (a code indicating "unknown" on HCFA's list), then the value is assigned to the unknown category. All other values are considered erroneous.

---

[5]There is an inconsistency in the treatment of birth year and graduation year in that dates of birth between 1900 and 1915 are not considered dubious, whereas graduation years consistent with such birth dates are considered erroneous or dubious.

[6]The carrier record sent to the Registry only allows for two digits for graduation year. The file we used, however, had a larger field size and thus values such as '1930' appeared in the data. Since they did not conform to HCFA's specifications, such values are considered erroneous.

[7]The sanction date field format is MMYY. If the month portion is unavailable, it appears as blanks in the data. If the year portion was feasible, then blanks in the month portion was considered feasible as well.

Feasible lengths of sanctions, according to HCFA's Office of Investigations (OIG), are values one (1) through 15, and starting with 15, in increments of five up through 50. A value of 99 indicates an indefinite sanction length and is also considered feasible. A value of zero indicates either a non-sanctioned provider or that the carrier did not know the sanction length.[8] Like sanction date, a zero-filled field is considered feasible. A value of blank indicates missing. Following OIG's suggestion, all non-feasible numeric values between 0 and 99 are considered dubious. Any other value not specified above is considered erroneous.

For the primary specialty code, only "UPIN" specialty codes are considered feasible (Exhibit 3-3). HCFA assigned the specialty code value 99 to "unknown;" we followed this practice and classified 99 into the unknown category. Zeroes and blanks are considered missing. All other specialty codes (values) are considered erroneous. Inspection of the data revealed that all of the erroneous values were equal to 49. Although it once indicated "miscellaneous" physician specialties, 49 now indicates ambulatory surgical centers. Since HCFA instructed carriers to revise specialty codes, occurrences of 49 are considered erroneous in this report. With one exception, secondary specialty codes were classified the same as primary specialty codes. Since the secondary specialty code is not required and because a provider may not have a secondary specialty, a zero-filled field (i.e., two zeroes), is not considered as missing.[9]

For state licensed in, only standard state postal codes are considered feasible. A blank indicates a missing value. All other values are considered erroneous. For state license number, a value equal to all blanks is considered missing. Since we do not know the correct format for each state's license number system, all other values are considered feasible.

---

[8]As with the date of death, HCFA's coding instructions made no provision for a situation in which a provider is known to be sanctioned but that carrier does not know any of the details. That *is*, for all three sanction data elements, blank- or zero-filled values were not intended to signify "unknown" data. Rather, only by default do we *infer* that to be the case.

[9]Note that in NCH data, a value of 00 in the CWFB HCFA Provider Specialty Code field indicates a clinical diagnosis laboratory fee screen.

The physician status data element has only two feasible values: 1 and 2. All other values are considered erroneous. Since neither of the board certification data elements are required for the Registry, values of Y (yes), N (no), U (unknown), and blank are all considered feasible values. All other values are considered erroneous.

There are only two feasible values for group practice indicator: 1 (group) and 4 (solo). Blanks are considered missing. All other values are considered erroneous. There are only two feasible values for the physician participation indicator: Y (yes) and N (no); all other values are considered erroneous.

3.3 Cross-Classification Variables

Values classified as missing, erroneous, dubious, or unknown were counted and percentage rates were calculated for each of these categories. These counts/rates are henceforth referred to as MEDU (for missing, erroneous, dubious, or unknown) counts/rates.

MEDU rates for each data element were calculated on several bases. First, unconditional MEDU rates were calculated nationally and by carrier. Next, national and carrier rates were calculated by a cross-classification (control) variable: DRIP status, the year the practice setting was added to the Registry, and by provider credential. Each of the control variables is described below as well as the reason for choosing it.

The type and/or activity of the practice setting is represented by the DRIP status.[10] The primary reason for stratifying on DRIP status is because the presence of inaccurate data on de-activated settings is of less concern than its presence on active settings. Of the nearly two million practice settings on the Registry file, about 317,000 are de-activated. As shown in Exhibit 3-4, there are 1,923 practice settings whose DRIP values do not conform to HCFA's

[10] In the Registry's documentation, the DRIP field is named MPIER-RES-INTERN-CODE and has the following coding: D - De-activated [practice setting], R - Resident, I - Intern, P- [active] Practice [setting].

EXHIBIT 3-4

FREQUENCIES OF CROSS-CLASSIFICATION VARIABLE VALUES

Value	Description	Frequency	Percent
DRIP Status			
P	Practice	1,663,321	83.20 %
D	De-activated	316,957	15.86
R	Resident	16,043	0.80
I	Intern	824	0.04
blank	no value present	1,922	0.10
	non-printable character	1	0.00
Year Setting Added to the Registry			
89	1989	969,755	48.51
90	1990	240,271	12.02
91	1991	319,804	16.00
92	1992	266,194	13.32
93	1993	203,043	10.16
	non-printable character	1	0.00
Provider Credentials			
MD	Medical Doctor	1,704,782	85.28
DO	Doctor of Osteopathy	67,300	3.37
DDS	Doctor of Dental Surgery	45,204	2.26
DDM	Doctor of Dental Medicine	11,098	0.56
CH	Chiropractor	73,309	3.67
DPM	Podiatrist	40,764	2.04
OD	Doctor of Optometry	56,610	2.83
	non-printable character	1	0.00

Note: There are 1,999,068 practice settings on the Registry file.

SOURCE: Physician Registry, July 1993.

specifications. The carriers with missing (no value present) or invalid DRIP values are shown in Exhibit 3-5. Kentucky BS had nearly 61 percent of the missing DRIP values, but Connecticut General Life (Idaho) had the highest rate. For our analysis, we created new categories for the DRIP status: (1) *active*, which included residents and interns as well as "practice" settings, (2) de-activated settings, and (3) missing, which included a case with a non-printable character (Exhibit 3-4).[11]

We also stratified our analyses by the year that setting was added to the Registry: (1) 1989, (2) 1990-1, and (3) 1992-3.[12] Given the implementation problems noted in the case studies, we expected that initially there would be relatively high MEDU rates, which would then decline as carriers and the Registry became more experienced with the procedures by which to submit data and to detect errors.

Finally, we stratified on the basis of credentials, that is, whether the provider is a physician (MD), doctor of osteopathy (DO), a dentist (doctor of dental medicine or dental surgery), or another limited license practitioner (podiatrist, optometrist, or chiropractor).[13] The last three limited license practitioners (LLPs) were grouped together because the differences in the MEDU rates among them were too small to justify analyzing separately. Except for one nonsensical value, all of the credential values are feasible; note, however, it was necessary to left-justify values before performing statistical analysis. One reason to analyze data integrity by credentials is that some types of providers might be more likely than others to cause information collection difficulties for carriers.

The analysis of *practice setting* data is straightforward: (1) each practice setting is counted once in all statistical procedures, and (2) all analyzed data elements come directly from

[11]There is one practice setting that has a non-printable character instead of a feasible year value (Exhibit 3-4). This practice setting is also associated with nonsensical values for several other data elements (including provider credentials) and therefore was omitted from all statistical analysis.

[12]The count for 1993 in Exhibit 3-4 represents just the first six months.

[13]HCFA considers dentists to be limited licensed practitioners. For this report, however, dentists were separated from the others.

EXHIBIT 3-5

CARRIERS WITH MISSING DRIP VALUES

Carrier Number	Area	Carrier Name	Missing DRIP	PERCENTAGE OF: All of the Carrier's Practice Settings	PERCENTAGE OF: All Problem Cases
00660	Kentucky*	Kentucky BS	1,170	5.49 %	60.84 %
00542	Northern California	California BS	336	0.21	17.47
05130	Idaho	Connecticut General Life	272	6.32	14.14
00740	Missouri (K.C.)	Kansas BS	86	0.74	4.47
02050	Southern California	TOLIC	27	0.02	1.40
00700	Massachusetts	Massachusetts BS	9	0.01	0.47
00510	Alabama	Alabama BC/BS	4	0.01	0.21
00880	South Carolina	South Carolina BC/BS	4	0.02	0.21
00520	Arkansas	Arkansas BC/BS	3	0.02	0.16
00528	Louisiana	Arkansas BC/BS	2	0.01	0.10
00550	Colorado	Colorado BC/BS	2	0.01	0.10
00655	Nebraska	Kansas BS	2	0.05	0.10
00720	Minnesota (Rural)	Minnesota BC/BS	2	0.01	0.10
00751	Montana	Montana BC/BS	2	0.05	0.10
05535	North Carolina	Connecticut General Life	1	0.00 **	0.05
10230	Connecticut	Travelers Ins. Co.	1	0.01	0.05
Totals:			1,923		100.00

*One observation with a non-printable character was classified as missing.

**Less than 0.01 percent.

SOURCE: Physician Registry, July 1993.

the practice setting section. On the other hand, because of the replication of *header* data across practice settings, only the header data from the first practice setting for a given provider is used in the statistical procedures. The interpretation of header MEDU rates by carrier needs to be tempered by the fact that more than one carrier may be contributing the information contained in the header section. Additionally, information on the cross-classification variables (DRIP, date setting added, and credentials) comes from the *practice setting* section of the UPIN's first record since header equivalents do not exist.

4.0 INTEGRITY ANALYSIS OF INDIVIDUAL DATA ELEMENTS

The integrity of the data elements listed in Exhibit 3-3 is examined in this chapter. The values of each data element are classified as feasible or missing/erroneous/dubious/unknown (MEDU) according to the criteria listed in Exhibit 3-3. The national MEDU frequency and percentage rates are presented and analyzed for each data element in the first section of this chapter. In the second section, we investigate the degree to which individual carriers have a disproportionate impact on national MEDU rates, for the four data elements that have the highest overall MEDU rates, by identifying the carriers with the highest and lowest MEDU rates. Individual summaries of each carrier's MEDU rates for each of the data elements listed in Exhibit 3-3 are presented in the third section.

4.1 National-Level Analysis

In this discussion, the focus is on data elements whose MEDU rate is 0.5 percent or higher.[1] The percentages reported in the tables allow detection of MEDU rates as low as 5 per 100,000 records. That is, if a carrier has 100,000 records, MEDU frequencies between zero and four (inclusive) would all appear as a percentage of zero, due to rounding. For most individual carriers, this means that the presence of any more than a couple of errors will be detected in nonzero percentages reported in the tables. For the national rates, however, this level of accuracy is not sufficient. Therefore, Table 4-1 contains, for all records, the actual MEDU frequency. Table 4-2 contains, for all records, the MEDU rates in percentage terms. Asterisks in Table 4-2 indicate the reported zero rate is rounded off from a nonzero rate smaller than 0.005 percent. At the top of each column in Table 4-2 and the carrier-specific tables in

[1] A MEDU rate of 0.5 percent is the same as 5 problems per 1,000 records.

TABLE 4-1

MEASURES OF VARIABLE INTEGRITY - NATIONAL FREQUENCIES

		DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
Variable	Overall	Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Header Variables											
Total Observations	693,355	584,219	107,607	1,529	465,463	139,088	88,804	530,438	22,259	46,296	94,362
Birth Year											
Dubious Values	78	43	35	0	71	3	4	58	2	5	13
Erroneous Values	1,586	1,340	240	6	909	380	297	1,079	79	111	317
Missing Values	62,910	52,582	10,328	0	56,960	5,280	670	59,117	565	285	2,943
Death Year											
Dubious Values	1,683	656	1,027	0	1,630	43	10	1,643	6	6	28
Erroneous Values	1	0	1	0	1	0	0	1	0	0	0
Missing Values	5	4	1	0	5	0	0	4	0	0	1
Medical School											
Erroneous Values	4,645	3,771	850	24	2,888	1,261	496	2,587	1,096	14	948
Missing Values	76,843	62,089	14,046	708	72,086	3,993	764	68,992	1,365	2,041	4,445
Grad Year											
Dubious Values	6,993	4,833	2,140	20	5,956	781	256	5,547	275	369	802
Erroneous Values	1,801	1,435	364	2	1,586	168	47	1,506	67	93	135
Missing Values	106,489	89,261	16,972	256	96,148	9,078	1,263	79,702	1,447	7,945	17,395
Year Sanctioned											
Erroneous Values	118	80	38	0	78	30	10	106	4	4	4
Missing Values	0	0	0	0	0	0	0	0	0	0	0
Sanction Code											
Unknown Values	230	211	19	0	216	9	5	216	4	0	10
Sanction Length											
Dubious Values	73	58	15	0	34	29	10	68	2	2	1
Missing Values	0	0	0	0	0	0	0	0	0	0	0

TABLE 4-1 (continued)

MEASURES OF VARIABLE INTEGRITY - NATIONAL FREQUENCIES

Variable	Overall	DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
		Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Setting Variables											
Total Observations	1,999,067	1,680,187	316,957	1,923	969,755	560,075	469,237	1,704,782	67,300	56,302	170,683
Birth Year											
Dubious Values	116	60	56	0	92	13	11	96	2	5	13
Erroneous Values	3,283	2,804	475	4	1,092	1,166	1,025	2,588	127	118	450
Missing Values	316,738	268,456	48,280	2	243,296	50,236	23,206	303,545	3,386	668	9,139
Death Year											
Dubious Values	2,028	683	1,345	0	1,930	59	39	1,988	6	5	29
Erroneous Values	1	0	1	0	1	0	0	1	0	0	0
Missing Values	5	4	1	0	4	0	1	4	0	0	1
Medical School											
Erroneous Values	11,664	9,600	2,056	8	5,399	3,592	2,673	6,487	3,520	19	1,638
Missing Values	361,521	293,477	66,595	1,449	289,967	46,604	24,950	339,528	7,852	2,868	11,273
Grad Year											
Dubious Values	10,216	7,198	2,998	20	7,619	1,697	900	8,466	359	399	992
Erroneous Values	4,364	3,923	439	2	1,268	2,953	143	4,001	84	102	177
Missing Values	441,215	365,613	75,199	403	351,361	61,272	28,582	386,567	7,808	9,876	36,964
Year Sanctioned											
Erroneous Values	180	135	45	0	93	70	17	163	2	6	9
Missing Values	0	0	0	0	0	0	0	0	0	0	0
Sanction Code											
Unknown Values	1,182	1,133	49	0	1,139	37	6	1,141	22	2	17
Sanction Length											
Dubious Values	98	82	16	0	11	69	18	95	0	2	1
Missing Values	0	0	0	0	0	0	0	0	0	0	0

TABLE 4-1 (continued)

MEASURES OF VARIABLE INTEGRITY - NATIONAL FREQUENCIES

Variable	Overall	DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
		Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Primary Specialty											
Erroneous Values	2,184	1,901	282	1	1,070	838	276	1,511	341	329	3
Missing Values	0	0	0	0	0	0	0	0	0	0	0
Unknown Values	629	532	97	0	164	290	175	297	22	305	5
Secondary Specialty											
Erroneous Values	53	38	15	0	48	5	0	52	1	0	0
Missing Values	1,884,816	1,582,592	300,302	1,922	916,383	522,382	446,051	1,604,045	62,872	54,815	163,084
Unknown Values	47	47	0	0	0	12	35	39	2	3	3
State Licensed In											
Erroneous Values	1	1	0	0	0	0	1	1	0	0	0
Missing Values	0	0	0	0	0	0	0	0	0	0	0
State License Number											
Missing Values	270,375	234,132	36,230	13	178,793	64,186	27,396	234,039	5,637	7,809	22,890
Physician Status											
Erroneous Values	1	1	0	0	0	0	1	1	0	0	0
Primary Board Certification											
Erroneous Values	0	0	0	0	0	0	0	0	0	0	0
Secondary Board Certification											
Erroneous Values	0	0	0	0	0	0	0	0	0	0	0
Group Practice Indicator											
Erroneous Values	10,197	8,314	1,869	14	5,634	3,168	1,395	8,226	364	580	1,027
Missing Values	0	0	0	0	0	0	0	0	0	0	0
Participation Indicator											
Erroneous Values	1	1	0	0	0	0	1	1	0	0	0

SOURCE: Physician Registry, July 1993.

TABLE 4-2

MEASURES OF VARIABLE INTEGRITY - NATIONAL RATES (PERCENTAGES)

		DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
Variable	Overall	Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Header Variables											
Total Observations	693,355	584,219	107,607	1,529	465,463	139,088	88,804	530,438	22,259	46,296	94,362
Threshold	35	30	6	1	24	7	5	27	2	3	5
Birth Year											
Dubious Values	0.01 %	0.01 %	0.03 %	0.00 %	0.02 %	0.00 %*	0.00 %*	0.01 %	0.01 %	0.01 %	0.01 %
Erroneous Values	0.23	0.23	0.22	0.39	0.20	0.27	0.33	0.20	0.35	0.24	0.34
Missing Values	9.07	9.00	9.60	0.00	12.24	3.80	0.75	11.14	2.54	0.62	3.12
Death Year											
Dubious Values	0.24	0.11	0.95	0.00	0.35	0.03	0.01	0.31	0.03	0.01	0.03
Erroneous Values	0.00 *	0.00	0.00 *	0.00	0.00 *	0.00	0.00	0.00 *	0.00	0.00	0.00
Missing Values	0.00 *	0.00 *	0.00 *	0.00	0.00 *	0.00	0.00	0.00 *	0.00	0.00	0.00 *
Medical School											
Erroneous Values	0.67	0.65	0.79	1.57	0.62	0.91	0.56	0.49	4.92	0.03	1.00
Missing Values	11.08	10.63	13.05	46.30	15.49	2.87	0.86	13.01	6.13	4.41	4.71
Grad Year											
Dubious Values	1.01	0.83	1.99	1.31	1.28	0.56	0.29	1.05	1.24	0.80	0.85
Erroneous Values	0.26	0.25	0.34	0.13	0.34	0.12	0.05	0.28	0.30	0.20	0.14
Missing Values	15.36	15.28	15.77	16.74	20.66	6.53	1.42	15.03	6.50	17.16	18.43
Year Sanctioned											
Erroneous Values	0.02	0.01	0.04	0.00	0.02	0.02	0.01	0.02	0.02	0.01	0.00
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sanction Code											
Unknown Values	0.03	0.04	0.02	0.00	0.05	0.01	0.01	0.04	0.02	0.00	0.01
Sanction Length											
Dubious Values	0.01	0.01	0.01	0.00	0.01	0.02	0.01	0.01	0.01	0.00 *	0.00 *
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 4-2 (continued)

MEASURES OF VARIABLE INTEGRITY - NATIONAL RATES (PERCENTAGES)

		DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
Variable	Overall	Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Setting Variables											
Total Observations	1,999,067	1,680,187	316,957	1,923	969,755	560,075	469,237	1,704,782	67,300	56,302	170,683
Threshold	100	85	16	1	49	29	24	86	4	3	9
Birth Year											
Dubious Values	0.01	0.00 *	0.02	0.00	0.01	0.00 *	0.00 *	0.01	0.00 *	0.01	0.01
Erroneous Values	0.16	0.17	0.15	0.21	0.11	0.21	0.22	0.15	0.19	0.21	0.26
Missing Values	15.84	15.98	15.23	0.10	25.09	8.97	4.95	17.81	5.03	1.19	5.35
Death Year											
Dubious Values	0.10	0.04	0.42	0.00	0.20	0.01	0.01	0.12	0.01	0.01	0.02
Erroneous Values	0.00 *	0.00	0.00 *	0.00	0.00 *	0.00	0.00	0.00 *	0.00	0.00	0.00
Missing Values	0.00 *	0.00 *	0.00 *	0.00	0.00 *	0.00	0.0 *	0.00 *	0.00	0.00	0.00 *
Medical School											
Erroneous Values	0.58	0.57	0.65	0.42	0.56	0.64	0.57	0.38	5.23	0.03	0.96
Missing Values	18.08	17.47	21.01	75.35	29.90	8.32	5.32	19.92	11.67	5.09	6.60
Grad Year											
Dubious Values	0.51	0.43	0.95	1.04	0.79	0.30	0.19	0.50	0.53	0.71	0.58
Erroneous Values	0.22	0.23	0.14	0.10	0.13	0.53	0.03	0.23	0.12	0.18	0.10
Missing Values	22.07	21.76	23.73	20.96	36.23	10.94	6.09	22.68	11.60	17.54	21.66
Year Sanctioned											
Erroneous Values	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Sanction Code											
Unknown Values	0.06	0.07	0.02	0.00	0.12	0.01	0.00 *	0.07	0.03	0.00 *	0.01
Sanction Length											
Dubious Values	0.00 *	0.00 *	0.01	0.00	0.00 *	0.01	0.00 *	0.01	0.00	0.00 *	0.00 *
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 4-2 (continued)

MEASURES OF VARIABLE INTEGRITY - NATIONAL RATES (PERCENTAGES)

		DRIP STATUS			DATE SETTING ADDED			CREDENTIALS			
Variable	Overall	Active	De-Activ.	Missing	1989	1990-1	1992-3	MD	DO	Dentists	Other LLP
Primary Specialty											
Erroneous Values	0.11	0.11	0.09	0.05	0.11	0.15	0.06	0.09	0.51	0.58	0.00 *
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Unknown Values	0.03	0.03	0.03	0.00	0.02	0.05	0.04	0.02	0.03	0.54	0.00 *
Secondary Specialty											
Erroneous Values	0.00 *	0.00 *	0.00 *	0.00	0.00 *	0.00 *	0.00	0.00 *	0.00 *	0.00	0.00
Unknown Values	0.00 *	0.00 *	0.00	0.00	0.00	0.00 *	0.01	0.00 *	0.00 *	0.01	0.00 *
State Licensed In											
Erroneous Values	0.00 *	0.00 *	0.00	0.00	0.00	0.00	0.00 *	0.00 *	0.00	0.00	0.00
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
State License Number											
Missing Values	13.53	13.93	11.43	0.68	18.44	11.46	5.84	13.73	8.38	13.87	13.41
Physician Status											
Erroneous Values	0.00 *	0.00 *	0.00	0.00	0.00	0.00	0.00 *	0.00 *	0.00	0.00	0.00
Primary Board Certification											
Erroneous Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Secondary Board Certification											
Erroneous Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Group Practice Indicator											
Erroneous Values	0.51	0.49	0.59	0.73	0.58	0.57	0.30	0.48	0.54	1.03	0.60
Missing Values	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Participation Indicator											
Erroneous Values	0.00 *	0.00 *	0.00	0.00	0.00	0.00	0.00 *	0.00 *	0.00	0.00	0.00

*Not true zeros

SOURCE: Physician Registry, July 1993

Appendix A are the total observations and the *threshold* value. (Separate sets of observation counts and threshold values are listed for the header versions and the setting versions of variables.) Frequencies less than the threshold value will result in a zero percentage being reported in the table. For example, Table A-8 (Appendix A), for the Florida carrier, indicates that the zero percent for dubious birth year values (setting version) could actually result from counts of either 0, 1, 2, 3, 4, or 5 dubious values because the threshold is 6. Rates having a value of zero in the tables and having a threshold value equal to one are true zeroes. Because of the large number of tables, we did not inspect each entry in the carrier tables to determine whether a zero was a true zero or not.

4.1.1 Header Data Elements

Of the header data elements listed in Exhibit 3-1, only the name variables are not analyzed in this section. None of the three sanction data elements, regardless of stratification, had MEDU rates equaling or exceeding 0.5 percent (Tables 4-1 and 4-2). The total count of non-blank sanction codes (not shown on the tables) suggests that only 1,635 providers were sanctioned out of 693,355 providers with UPINs assigned. Of these 1,635 providers, the unknown code was reported for 230, or 14 percent of those sanctioned. There were also 118 erroneous values for the year sanctioned value. Although there were not any erroneous values reported for sanction length, 73 of the values were classified as dubious because they did not conform to OIG specifications (see section 3.2.2).

Three of the four other header data elements analyzed had notable MEDU rates, especially due to missing values. About nine percent of the header dates of birth were missing. The difference in rates of missing values by active versus inactive status is not large. The rates of missing birth year, however, have declined over time from about 12 percent for practice settings added to the Registry in 1989 to less than one percent for those added in 1992-3.

Missing birth year rates for all four credential strata exceeded 0.5 percent, with that for MDs being the worst at 11 percent.

The rates of missing values for the header medical school codes have similar patterns to those for birth year, but at slightly higher levels. For instance, the missing value rates for both birth year and medical school declined over time. It should be noted, however, that medical school was not a required data element until December 1991. The error rates for medical school code usually range between 0.5 and 1.0 percent. The error rate for 1992-93 is only slightly lower than that in 1989; and the error rate is actually highest during 1990-91. However, the changes in error rates should be considered in conjunction with the dramatic declines in the missing value rate. The medical school error rate for DOs is especially high, at nearly 5 percent.

For graduation year, the missing rates were quite high; and, as with medical school codes, the missing rates decline over time. As with medical school, graduation year was not a required data element until December 1991. The rates of dubious values for graduation year range between 0.29 and 1.28 percent. The rate of dubious values declines considerably over time. The rate of dubious values is high for MDs, although it is highest for the DOs.

4.1.2 Setting Data Elements

The setting versions of birth year, medical school, and graduation year also have high MEDU rates, primarily due to missing values. They exhibit patterns similar to their header counterparts, albeit at higher levels. It is not clear why MEDU rates are higher for practice setting versions of data elements than for their header counterparts. One possible reason is that header MEDU rates improve relative to setting MEDU rates because of TOLIC header construction practices. Since TOLIC does not replace feasible header values by non-feasible values when a new practice setting is added or an update record is submitted, only one of the

provider's practice setting need have feasible values for the header value to be feasible. On the other hand, if there are multiple practice settings, it only takes one practice setting with non-feasible values to generate a MEDU value. For instance, while one carrier might have submitted a feasible birth year value for a given provider, another carrier might have submitted a practice setting that did not have a birth year recorded (i.e., birth year was missing). Thus, for this hypothetical provider, the header value of birth year is feasible, but one-half of the practice settings have a missing birth year. As demonstrated in Chapter 6 some carriers have inconsistencies among their own records. The most likely explanation for why MEDU rates are higher for practice setting values than for header values, however, is that there may be multiple setting records for a provider, and TOLIC has the ability to "fill in" header values from any one of those settings.

Of the other setting variables, primary specialty, license number, and group practice indicator have MEDU rates exceeding 0.5 percent. For primary specialty, error rates for DOs and dentists were high. For license numbers, the problem is missing values. We were not able to check for erroneous or dubious values. Since license numbers were not required during the implementation phase of the Registry, missing value rates have declined over time. The errors found in the group practice indicator, which have declined over time, may be due to some carriers' use of an older, more comprehensive coding scheme that has values ranging from 1 to 4, while the Registry allows only codes 1 and 4, and not codes 2 or 3.

4.2 Carriers with Highest and Lowest MEDU Rates

In the previous section it was shown that the data elements with the highest combined MEDU rates were: (1) birth year, (2) school code, (3) graduation year, and (4) state license number. The combined MEDU rates of each of the four data elements fell continuously after

the implementation phase in 1989. Additionally, it was found that the combined MEDU rates were often highest for medical doctors.

The prior analysis was conducted at the national level without taking into regard the individual carriers. There may have been, however, some carriers that had more difficulty than others in obtaining or submitting accurate information. We want to know, then, whether some carriers had especially high MEDU rates while others had very low rates. For instance, did the same carriers that had the highest MEDU rates during the implementation phase also have the highest rates in 1992-93? Similarly, did the same carriers have difficulties obtaining information from some types of providers but not others? Did some carriers have relatively high MEDU rates for some data elements but not others?

To help determine whether some carriers had consistently high or low MEDU rates, for each of these four data elements, the carriers with the highest and lowest MEDU rates are identified. Since the values of the header versions of birth year, school code, and graduation year can be obtained from more than one carrier, only the practice setting versions are examined. Of the four data elements, only birth year was a mandatory data element during the implementation phase; the other three became mandatory in December 1991.

Most carriers submitted over half of their eventual (by mid-1993) practice settings during 1989. Thus their 1989 MEDU rates substantially influence their overall MEDU rates. Unlike other carriers, however, the two Railroad Retirement Board (RRB) carriers did not participate in the first part of the implementation phase. The Georgia RRB submitted only two practice settings in 1989 while the Utah RRB submitted 25,157 in 1989 (out of the 143,270 it had by mid-1993). Therefore, their 1989 MEDU rates should not be taken as indicative of their overall rates.

It should be noted that the individual carrier denominators do not sum to the total number of providers serviced by the carrier, due to the existence of multiple practice settings

for many providers. Similarly, carrier provider counts would not sum to the national total due to some providers' multiple carrier affiliations.

4.2.1 Birth Year

The four carriers with the highest overall combined MEDU[2] rates for birth year were Aetna (Georgia), C & S Administrative Services (Massachusetts), Arkansas BC/BS (Arkansas), and South Carolina BC/BS (Table 4-3) with rates 40 percent and higher. With only eight practice settings, only S.S.S. (Virgin Islands) had no errors or missing values. California BS (northern California), North Dakota BC/BS (Wyoming), and Washington Physician Service had the next three lowest MEDU rates – all under 0.5 percent.

Aside from the two RRB carriers, the four carriers with the highest MEDU rates in 1989 were Aetna (Georgia and New Mexico), Minnesota BC/BS, and Arkansas BC/BS (Arkansas), all with rates 64 percent and higher. There were another four carriers with combined MEDU rates exceeding 50 percent. The four carriers with lowest rates in 1989 were S.S.S. (Virgin Islands), California BS (northern California), North Dakota BC/BS (Wyoming), and S.S.S. (Puerto Rico). Only S.S.S. (Virgin Islands) and California BS had rates less than 0.5 percent.

In 1990-91, the two RRB carriers had the highest MEDU rates with 33.2 percent for the Georgia operations and 22.2 percent for the Utah operations. As previously mentioned, the two RRB carriers did not submit most of their *eventual* practice settings during 1989. The 33.2 percent value for the Georgia operations would have, *ceteris paribus*, placed it as the 22nd highest carrier during 1989. Besides the two RRB carriers, only two other carriers had MEDU rates greater than the national average of nine percent: Travelers (Connecticut) and Arkansas BC/BS (Arkansas).

[2]For birth year, the relevant MEDU components are missing, erroneous, and dubious. The word *combined* refers to the sum of the three components. (Even when *MEDU* stands alone, it should be understood that the combined rate is inferred.) The word *overall* refers to combined MEDU rate regardless of the year setting added to the Registry and regardless of provider type.

TABLE 4-3

BIRTH YEAR (SETTING VERSION), COMBINED MISSING/ERRONEOUS/DUBIOUS PERCENTAGE RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Carrier Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added* 1989	1990-1	1992-3	*Provider Type* MD	DO	Dentist	Other LLP
	National		1,999,067	16.01 %	25.21 %	9.18 %	5.17 %	17.97 %	5.22 %	1.41 %	5.62 %
01040	Aetna Life & Casualty	GA	27,908	54.52	89.46	5.33	0.46	60.35	12.27	0.26	0.10
00700	C&S Admin. Services	MA	62,426	40.57	57.34	8.21	0.34	45.70	24.93	0.23	0.37
00520	Arkansas BC/BS	AR	15,821	39.83	64.26	10.18	0.11	44.91	11.37	0.17	0.00
00880	South Carolina BC/BS	SC	24,095	39.44	54.17	1.33	0.31	45.10	0.00	0.00	0.00
00630	Indiana BC/BS	IN	28,863	33.50	58.11	0.36	0.44	37.47	0.58	0.21	0.20
00720	Minnesota BC/BS	MN	14,907	31.14	64.38	0.53	0.18	38.05	7.61	0.00	0.18
00781	C&S Admin. Services	VT	4,585	30.52	45.97	0.57	0.44	35.26	6.82	0.00	0.00
21200	C&S Admin. Services	ME	6,631	30.37	45.92	3.78	1.41	38.32	2.31	0.70	1.03
00803	Empire BC/BS	NY	92,461	30.16	39.07	0.13	1.10	34.18	15.74	0.12	0.31
01360	Aetna Life & Casualty	NM	7,337	29.75	66.86	1.04	0.23	33.82	11.59	1.65	4.69
10074	Travelers-RRB-UT	RRB	143,270	29.22	71.80	22.23	19.21	29.71	65.55	25.36	23.93
10230	Travelers Ins. Co.	CT	17,682	27.38	41.54	16.22	0.79	31.27	0.00	0.42	0.79
00910	Utah BC/BS	UT	6,432	26.07	46.23	3.11	0.53	29.73	0.00	1.23	3.21
00621	Illinois BC/BS	IL	70,990	25.45	43.31	1.33	0.33	29.40	0.93	0.50	0.48
01370	Aetna Life & Casualty	OK	13,152	25.44	47.44	1.90	0.26	32.75	6.21	6.84	3.90
01030	Aetna Life & Casualty	AZ	20,462	24.34	57.45	0.99	0.29	30.34	4.74	0.00	0.21
10072	Travelers-RRB-GA	RRB	183,394	24.28	100.00	33.16	10.58	23.89	49.54	22.26	27.09
00570	Pennsylvania BS	DE	3,557	22.97	35.44	2.88	0.23	27.60	10.47	0.00	0.77
00580	Pennsylvania BS	DC	20,151	21.31	33.20	4.27	0.71	23.22	10.56	0.00	0.37
00690	Maryland BS	MD	25,740	20.98	39.11	0.56	0.33	22.80	0.00	0.38	0.27
05130	Connecticut General	ID	4,307	20.10	31.21	0.37	0.40	24.18	0.68	0.00	0.20
01120	Aetna Life & Casualty	HI	5,982	19.91	43.69	0.31	0.40	22.08	4.65	1.45	0.64
00900	Texas BC/BS	TX	96,351	19.05	36.20	0.62	0.21	23.33	2.08	0.21	0.38
14330	Group Health Ins., Inc.	NY	9,609	18.31	40.74	0.48	0.05	21.60	0.64	0.00	0.18
00801	BS of Western NY	NY	28,702	16.91	31.78	0.30	0.29	18.85	21.50	0.15	0.25
00865	Pennsylvania BS	PA	83,361	16.75	25.91	4.33	0.86	21.24	10.71	0.19	0.71
00590	Florida BC/BS	FL	106,031	14.38	24.16	1.40	0.44	17.16	3.66	0.27	0.22
00780	C&S Admin. Services	NH	6,645	12.33	25.21	1.27	0.74	14.24	3.08	0.92	0.87
10490	Travelers Ins. Co.	VA	22,058	9.13	15.11	1.97	0.13	10.09	10.00	0.55	0.19

TABLE 4-3 (continued)

BIRTH YEAR (SETTING VERSION), COMBINED MISSING/ERRONEOUS/DUBIOUS PERCENTAGE RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added* 1989	1990-1	1992-3	*Provider Type* MD	DO	Dentist	Other LLP
00528	Arkansas BC/BS	LA	30,248	8.58 %	15.22 %	0.49 %	0.15 %	9.17 %	3.51 %	0.00 %	0.20 %
10250	Travelers Ins. Co.	MS	11,209	8.32	13.87	0.71	0.16	8.90	20.00	0.00	0.18
00751	Montana BC/BS	MT	3,739	7.84	14.73	0.16	0.37	9.56	0.00	0.00	0.44
00820	North Dakota BC/BS	SD-ND	7,501	7.79	13.98	0.22	0.09	9.33	6.71	0.90	0.00
16510	Nationwide Ins. Co.	WV	15,163	7.61	12.34	0.98	0.17	8.60	3.35	0.80	0.11
00860	Pennsylvania BS	NJ	37,087	7.56	13.90	0.76	0.42	8.99	4.65	0.14	0.43
16360	Nationwide Ins. Co.	OH	84,069	6.33	11.16	0.55	0.18	7.69	2.58	0.23	0.04
02050	Transamerica Occidental	So. CA	117,191	5.29	9.35	0.54	0.27	5.64	11.58	1.45	0.59
05535	Connecticut General	NC	40,046	4.24	6.90	0.27	0.15	4.70	0.00	0.25	0.08
00655	Kansas BS	NE	7,045	3.66	10.81	0.51	0.00	4.29	2.41	0.00	0.00
00650	Kansas BS	KS	9,445	3.39	8.20	0.03	0.09	4.22	4.06	0.00	0.08
00510	Alabama BC/BS	AL	28,724	2.88	4.96	0.14	0.09	3.07	3.53	0.47	0.00
01290	Aetna Life & Casualty	NV	6,008	2.63	6.55	0.40	0.08	2. 0	3.07	0.00	0.00
10240	Travelers Ins. Co.	MN	15,323	2.12	3.58	0.31	0.42	2.41	10.34	0.00	0.06
01380	Aetna Life & Casualty	OR	17,800	2.03	3.54	0.21	0.18	2.13	1.89	4.85	0.83
00550	Colorado BC/BS	CO	16,308	1.79	4.96	0.25	0.07	2.02	1.47	0.00	0.14
01020	Aetna Life & Casualty	AK	2,382	1.51	3.69	0.24	0.00	1.77	0.82	0.00	0.00
00951	Wisconsin Physician Serv.	WI	24,456	1.49	1.95	1.55	0.10	1.83	0.85	0.16	0.00
11260	Missouri Gen. Amer. Life	MO	23,668	1.48	2.92	0.78	0.17	1.72	1.02	0.00	0.47
00870	Rhode Island BS	RI	5,925	1.43	1.76	0.89	1.30	1.81	0.00	0.00	0.00
00710	Michigan BC/BS	MI	66,214	1.40	3.04	0.50	0.27	1.71	0.81	0.60	0.16
00660	Kentucky BC/BS	KY	21,296	1.07	1.41	0.48	0.38	1.14	2.17	0.00	0.16
00740	Kansas BS	MO	11,636	1.05	1.91	0.19	0.05	1.29	0.09	0.00	0.17
00640	Iowa BS	IA	17,589	0.91	1.10	0.96	0.14	1.10	0.46	0.43	0.37
05440	Connecticut General	TN	26,944	0.75	0.93	0.41	0.40	0.79	0.00	0.00	0.60
00973	S.S.S.	PR	10,987	0.67	0.72	0.80	0.14	0.67	0.00	1.00	0.45
00932	Washington Physician Serv.	WA	23,763	0.41	0.78	0.04	0.11	0.47	0.15	0.00	0.13
00825	North Dakota BC/BS	WY	2,315	0.39	0.56	0.50	0.00	0.25	0.00	2.78	1.20
00542	California BS	No. CA	162,068	0.20	0.13	0.26	0.33	0.18	0.10	0.31	0.39
00974	S.S.S.	VI	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

SOURCE: Physician Registry, July 1993

For 1992-93, the two RRB carriers had the highest MEDU rates with 19.2 percent for the Utah operations and 10.6 percent for the Georgia operations. The two RRB MEDU rates were the only rates higher than the national average of 5.17 percent. The next two highest rates were for C & S Administrative Services (Maine) and Rhode Island BS, both under 1.5 percent. The mean of the non-RRB carriers was 0.33 percent with four carriers not having any errors or missing values: S.S.S. (Virgin Islands), Aetna (Alaska), North Dakota BC/BS (Wyoming), and Kansas BS (Nebraska).

The carriers with the highest and lowest MEDU rates for medical doctors were the same as for the overall rate. For doctors of osteopathy (DO), dentists, and other limited licensed practitioners, the two RRB carriers had the highest MEDU rates. For each of these last three provider categories, the next two highest MEDU rates were considerably lower – sometimes only one-fourth the RRB values. Eleven carriers had MEDU rates equal to zero for doctors of osteopathy and other limited licensed practitioners, while, for dentists, 24 carriers had MEDU rates equal to zero.

4.2.2 School Code

Despite not being a mandatory data element during the implementation phase, school code had only a slightly higher combined MEDU[3] rate than birth year. The same four carriers that had the highest overall MEDU rates for birth year were also the highest for school code, although with a different ranking. Aetna (Georgia) had 54.6 percent MEDU rate for school code while C & S Administrative Services, the fourth highest, had a 37.9 percent MEDU rate (Table 4-4). California BS (northern California) and S.S.S. (Puerto Rico) were the only two carriers with MEDU rates less than 0.5 percent. The next lowest was Wisconsin Physician Service at 1.6 percent followed by Missouri General and Aetna (Nevada) at 3.8 percent.

[3]For school code, the relevant components of MEDU are missing and erroneous.

TABLE 4-4

SCHOOL CODE (SETTING VERSION), COMBINED MISSING/ERRONEOUS (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Carrier Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added* 1989	1990-1	1992-3	*Provider Type* MD	DO	Dentist	Other LLP
	National		1,999,067	18.66 %	30.46 %	8.96 %	5.89 %	20.30 %	16.90 %	5.12 %	7.56 %
01040	Aetna Life and Casualty	GA	27,908	54.64	89.58	5.38	0.66	60.42	13.07	0.00	0.84
00880	South Carolina BC/BS	SC	24,095	46.53	63.82	1.49	0.93	45.09	0.00	71.94	45.66
00520	Arkansas BC/BS	AR	15,821	40.76	65.08	11.49	0.78	45.86	13.71	0.17	0.71
00700	C&S Administrative Services	MA	62,426	37.91	53.90	6.56	0.13	42.24	26.02	2.06	6.12
00640	Iowa BS	IA	17,589	36.55	61.46	4.94	1.44	46.39	23.98	0.71	1.84
01360	Aetna Life and Casualty	NM	7,337	35.33	78.78	2.08	0.28	33.72	19.82	66.12	53.13
01370	Aetna Life and Casualty	OK	13,152	32.92	61.69	2.01	0.14	31.58	6.15	69.94	72.95
21200	C&S Administrative Services	ME	6,631	32.15	48.87	3.93	0.51	37.40	9.59	29.47	8.42
00720	Minnesota BC/BS	MN	14,907	31.11	64.42	0.24	0.35	37.95	9.14	0.11	0.30
00630	Indiana BC/BS	IN	28,863	30.73	53.14	0.69	0.41	31.49	2.92	8.32	28.94
00781	C&S Administrative Services	VT	4,585	30.62	46.20	0.68	0.00	35.03	11.36	2.02	1.61
00803	Empire BC/BS	NY	92,461	30.30	39.25	0.33	0.87	34.34	24.26	0.00	0.02
00820	North Dakota BC/BS	ND-SD	7,501	29.76	50.44	5.82	1.45	29.66	22.56	32.13	29.79
10230	Travelers Insurance Co.	CT	17,682	29.25	52.88	4.49	0.55	30.84	36.67	15.65	19.25
01030	Aetna Life and Casualty	AZ	20,462	29.15	66.26	3.30	1.41	31.23	30.90	15.13	15.94
00932	Washington Physician Service	WA	23,763	28.79	59.00	1.54	1.67	26.95	6.78	53.81	41.64
00621	Illinois BC/BS	IL	70,990	28.44	45.75	5.08	3.95	30.81	52.88	0.69	0.95
16360	Nationwide Insurance Co.	OH	84,069	28.41	49.66	4.02	0.33	34.03	15.02	0.11	0.63
00910	Utah BC/BS	UT	6,432	27.89	50.70	1.22	0.09	32.19	1.89	0.61	0.15
10074	R.R.B. - Travelers - UT	RRB	143,270	27.47	72.70	11.69	20.64	30.68	67.07	4.36	3.94
05130	Connecticut General Life	ID	4,307	27.14	42.20	0.50	0.27	24.13	17.81	45.68	48.13
10072	R.R.B. - Travelers - GA	RRB	183,394	24.18	100.00	32.16	11.83	26.29	53.70	4.01	5.63
00900	Texas BC/BS	TX	96,351	23.14	42.86	2.19	0.99	24.62	17.20	0.94	28.02
00570	Pennsylvania BS	DE	3,557	22.43	34.59	2.89	0.23	26.93	10.83	0.58	0.00
00690	Maryland Blue Shield	MD	25,740	21.34	39.73	0.79	0.14	23.21	0.52	0.00	0.45
00580	Pennsylvania BS	DC	20,151	20.94	32.55	4.21	0.92	22.70	13.04	0.38	1.71
16510	Nationwide Insurance	WV	15,163	19.98	32.61	1.37	0.69	22.41	7.98	0.00	3.91
00801	Blue Shield of Western New York	NY	28,702	17.47	32.37	0.95	0.63	19.48	23.19	0.00	0.17
01020	Aetna Life and Casualty	AK	2,382	17.13	43.93	0.37	0.00	14.82	12.30	50.00	33.60

TABLE 4-4 (continued)

SCHOOL CODE (SETTING VERSION), COMBINED MISSING/ERRONEOUS (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Carrier Name	State	Number of Practice Settings	Overall Rate	Year Setting Added			Provider Type			
					1989	1990-1	1992-3	MD	DO	Dentist	Other LLP
00865	Pennsylvania BS	PA	83,361	16.36 %	25.40 %	4.14 %	0.60 %	20.72 %	10.64 %	0.55 %	0.27 %
00660	Kentucky BC/BS	KY	21,296	13.27	19.19	2.59	1.76	8.05	16.52	71.70	75.75
00710	Michigan BC/BS	MI	66,214	12.36	26.37	6.25	1.47	13.08	19.01	0.00	0.30
00780	C&S Administrative Services	NH	6,645	12.22	25.41	1.12	0.07	14.13	13.85	0.00	0.00
10250	Travelers Insurance Co.	MS	11,209	11.47	18.58	1.53	1.35	12.10	30.00	0.00	3.68
14330	Group Health Insurance, Inc.	NY	9,609	11.30	24.56	0.80	0.46	13.00	5.10	0.00	1.84
01120	Aetna Life and Casualty	HI	5,982	11.13	24.29	0.44	0.00	12.24	6.98	0.00	0.85
00860	Pennsylvania BS	NJ	37,087	10.22	18.47	1.55	0.52	11.67	12.27	0.00	0.47
10490	Travelers Insurance Co.	VA	22,058	9.62	15.90	1.95	0.34	10.33	10.00	0.55	3.83
00528	Arkansas BC/BS	LA	30,248	8.83	15.39	0.62	0.68	9.42	3.51	0.22	0.52
00740	Kansas BS	MO-KS	11,636	8.68	15.89	1.36	0.73	7.45	0.37	1.64	26.60
00650	Kansas BS	KS	9,445	7.96	18.12	1.01	0.78	4.95	5.59	4.35	28.54
00751	Montana BC/BS	MT	3,739	7.89	14.22	1.13	0.37	9.58	1.00	0.00	0.44
01380	Aetna Life and Casualty	OR	17,800	7.76	13.76	0.60	0.50	2.62	10.30	46.37	41.10
00825	North Dakota BC/BS	WY	2,315	7.69	15.07	6.53	0.99	7.54	3.80	0.00	11.16
00550	Colorado BC/BS	CO	16,308	7.61	20.23	1.67	0.45	8.35	9.39	0.00	0.29
02050	T.O.L.I.C.	So. CA	117,191	7.25	12.48	1.14	0.83	7.62	28.19	0.13	0.35
10240	Travelers Insurance Co.	MN	15,323	7.25	12.79	0.67	0.45	6.08	62.07	8.93	15.47
00590	Florida BC/BS	FL	106,031	6.65	10.35	2.13	0.80	7.42	9.50	0.02	0.78
00510	Alabama BC/BS	AL	28,724	6.43	11.06	0.34	0.09	6.86	5.87	0.47	0.49
00655	Kansas BS	NE	7,045	6.34	17.63	1.12	1.13	4.74	2.41	3.51	20.08
00870	Rhode Island Blue Shield	RI	5,925	4.81	6.15	5.04	1.59	4.53	34.48	0.43	10.65
05535	Connecticut General Life	NC	40,046	4.52	7.35	0.22	0.29	4.92	0.00	0.13	1.49
05440	Connecticut General Life	TN	26,944	4.09	6.16	0.34	0.13	3.45	4.63	0.71	15.99
01290	Aetna Life and Casualty, NV	NV	6,008	3.78	7.71	1.75	0.80	3.53	11.79	0.00	0.19
11260	Missouri General Am. Life	MO	23,668	3.77	6.88	2.38	0.64	2.01	18.24	1.08	1.57
00951	Wisconsin Physician Service	WI	24,456	1.59	1.83	2.04	0.48	1.38	13.72	0.00	0.85
00973	S.S.S.	PR	10,987	0.34	0.50	0.13	0.00	0.12	0.00	1.99	9.09
00542	California BS	No. CA	162,068	0.23	0.33	0.17	0.03	0.16	6.54	0.00	0.08

SOURCE Physician Registry, July 1993

Aside from the two RRB carriers, three of the four carriers with the highest combined MEDU rates in 1989 for birth year were the same for school code with Aetna (Arizona) replacing Minnesota BC/BS. The combined MEDU rates for each of the four highest were 65 percent and higher. There were another ten carriers with combined MEDU rates exceeding 50 percent. California BS (northern California) and S.S.S. (Puerto Rico) were the only two carriers with MEDU rates less than 0.5 percent. The next lowest was Wisconsin Physician Service at 1.8 percent followed by Rhode Island BS at 6.2 percent.

In 1990-91, the two RRB carriers had the highest MEDU rates with 32.2 percent for the Georgia operations and 11.7 percent for the Utah operations. The 32.2 percent value for the Georgia operations would have, *ceteris paribus*, placed it as the 28th highest carrier during 1989. Besides the two RRB carriers, only Arkansas BC/BS (Arkansas) had MEDU rates greater than the national average of 8.96 percent. The carrier with fourth highest rate was C & S Administrative Service (Massachusetts) at 6.56 percent followed closely by North Dakota BC/BS (Wyoming) and Michigan BC/BS. Joining California BS (northern California) and S.S.S. (Puerto Rico) with the lowest MEDU rates were Connecticut General Life (North Carolina) and Minnesota BC/BS, all with MEDU rates 0.2 percent or less.

For 1992-93, the two RRB carriers had the highest MEDU rates with 20.6 percent for the Utah operations and 11.8 percent for the Georgia operations. The two RRB MEDU rates were the only rates higher than the national average of 5.89 percent. The next two highest rates were for Illinois BC/BS at 3.95 percent and Kentucky BC/BS at 1.76 percent. There were 25 carriers with combined MEDU rates less than 0.5 percent, with four carriers having a rate equal to zero: S.S.S. (Puerto Rico), Aetna (Alaska and Hawaii), and C & S Administrative Services (Vermont).

The carriers with the highest and lowest MEDU rates for medical doctors were the same as for the overall rate with two exceptions: Iowa BS replaced C & S Administrative (Massachusetts) among the four highest and Aetna (Nevada) dropped out of the very lowest group. For doctors of osteopathy, the carriers with the highest MEDU rates, all exceeding 52.9

percent, were RRB (Utah), Travelers (Minnesota), RRB (Georgia), and Illinois BS. Three carriers had MEDU rates equal to zero for doctors of osteopathy: S.S.S. (Puerto Rico), Connecticut General Life (North Carolina), and South Carolina BC/BS. For dentists, the MEDU rates were quite high with four carriers having rates 66.1 percent and higher: South Carolina BC/BS, Kentucky BC/BS, and Aetna (Oklahoma and New Mexico). There were 17 carriers with MEDU rates equal to zero for dentists. For other limited license practitioners, Kentucky BC/BS had a combined MEDU rate of 75.8 percent, Aetna (Oklahoma) was at 72.95 percent, Aetna (New Mexico) was at 53.1 percent, Connecticut General Life (Idaho) was at 48.1 percent. C & S Administrative Services (New Hampshire) and Pennsylvania BS (Delaware) had MEDU rates equal to zero. The carriers with next lowest MEDU rates for other limited license practitioners were Empire BC/BS at 0.02 percent and California BS at 0.08 percent.

4.2.3 Graduation Year

Like school code, graduation year was not a mandatory data element during the implementation phase. Unlike school code, however, the overall combined MEDU[4] rate for graduation year was considerably higher than the combined MEDU rate for birth year. For graduation year, Wisconsin Physician Service had the highest overall MEDU rate at 67.1 percent, followed by Aetna (Georgia) at 54.8 percent, South Carolina BC/BS at 47.3 percent, and TOLIC at 44.2 percent (Table 4-5). The lowest overall MEDU rates ranged from 1.21 percent to 2.14 percent; the carriers were: S.S.S. (Puerto Rico), California BS, Connecticut General Life (Tennessee), and Iowa BS.

Aside from the Georgia RRB carrier, the highest MEDU rates in 1989 were for Wisconsin Physician Service at 98.7 percent, Aetna (Georgia) at 89.8 percent, Minnesota BC/BS at 86.9 percent, and Aetna (New Mexico) at 79.4 percent. With MEDU rates ranging from 1.7

[4]For graduation year, the components of MEDU are missing, erroneous, and dubious.

TABLE 4-5

GRADUATION YEAR (SETTING VERSION), COMBINED MISSING/ERRONEOUS/DUBIOUS (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Carrier Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added* 1989	1990-1	1992-3	*Provider Type* MD	DO	Dentist	Other LLP
	National		1,999,067	22.80 %	37.15 %	11.77 %	6.31 %	23.41 %	12.25 %	18.43 %	22.34 %
00951	Wisconsin Physician Service	WI	24,456	67.07	98.67	29.39	15.28	66.15	53.89	95.42	64.08
01040	Aetna Life and Casualty	GA	27,908	54.78	89.75	5.61	0.58	60.45	13.07	1.05	2.13
00880	South Carolina BC/BS	SC	24,095	47.33	64.98	1.78	0.28	46.10	0.00	72.14	44.10
02050	T.O.L.I.C.	So. CA	117,191	44.19	68.18	21.94	9.04	42.78	28.65	52.64	61.78
00720	Minnesota BC/BS	MN	14,907	41.89	86.93	0.22	0.21	37.86	8.12	81.98	53.56
00520	Arkansas BC/BS	AR	15,821	41.79	67.38	10.66	0.26	45.58	12.04	12.99	14.23
00700	C&S Administrative Services	MA	62,426	38.53	54.67	6.98	0.30	42.71	27.64	3.36	8.49
00820	North Dakota BC/BS	ND-SD	7,501	38.41	59.29	17.20	3.72	42.01	31.10	26.24	20.65
01360	Aetna Life and Casualty	NM	7,337	35.93	79.43	2.82	0.61	34.04	21.04	71.07	55.31
00781	C&S Administrative Services	VT	4,585	35.62	53.37	1.47	0.74	36.02	12.50	43.77	27.71
01370	Aetna Life and Casualty	OK	13,152	33.89	63.15	2.79	0.11	32.25	9.05	72.32	72.95
21200	C&S Administrative Services	ME	6,631	33.86	50.94	5.71	0.51	38.68	8.70	36.14	13.92
00910	Utah BC/BS	UT	6,432	33.05	59.66	2.33	0.00	33.56	5.66	25.77	32.77
16360	Nationwide Insurance Co.	OH	84,069	32.91	57.57	4.23	0.66	34.39	13.32	67.75	33.59
10230	Travelers Insurance Co.	CT	17,682	31.83	57.60	4.69	0.92	26.97	33.33	78.95	59.59
00621	Illinois BC/BS	IL	70,990	31.08	52.67	2.03	0.42	29.89	1.32	64.59	48.28
00803	Empire BC/BS	NY	92,461	30.43	39.31	0.59	1.37	34.25	17.87	2.25	1.92
10074	R.R.B. - Travelers - UT	RRB	143,270	29.78	72.30	22.48	19.92	30.22	65.45	26.07	24.87
00932	Washington Physician Service	WA	23,763	29.29	59.73	1.64	2.17	27.29	6.63	56.08	43.00
01030	Aetna Life and Casualty	AZ	20,462	27.86	65.29	1.52	0.49	30.71	25.96	31.79	8.86
10072	R.R.B. - Travelers - GA	RRB	183,394	26.60	100.00	35.93	12.18	26.39	53.70	24.83	27.64
00900	Texas BC/BS	TX	96,351	24.91	46.12	2.82	0.22	26.53	23.93	3.40	23.80
05535	Connecticut General Life	NC	40,046	24.68	39.16	4.48	0.76	25.10	54.55	13.53	25.62
00570	Pennsylvania BS	DE	3,557	23.64	36.43	3.11	0.23	28.13	10.47	2.31	2.70
00690	Maryland Blue Shield	MD	25,740	22.45	41.75	0.85	0.21	24.31	0.00	1.27	1.62
00630	Indiana BC/BS	IN	28,863	22.30	38.46	0.66	0.41	23.16	1.75	7.68	17.51
00801	Blue Shield of Western New York	NY	28,702	22.01	37.20	7.81	0.61	22.27	20.29	53.90	10.68
00580	Pennsylvania BS	DC	20,151	21.84	33.98	4.65	0.41	23.56	13.04	1.88	3.04
05130	Connecticut General Life	ID	4,307	20.41	31.64	0.50	0.40	24.52	0.68	0.00	0.39
00780	C&S Administrative Services	NH	6,645	17.89	35.28	3.80	1.14	16.17	10.77	34.48	24.46

TABLE 4-5 (continued)

GRADUATION YEAR (SETTING VERSION), COMBINED MISSING/ERRONEOUS/DUBIOUS (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Carrier Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added*			*Provider Type*			
					1989	1990-1	1992-3	MD	DO	Dentist	Other LLP
00865	Pennsylvania BS	PA	83,361	17.55 %	26.88 %	5.27 %	0.57 %	21.71 %	11.08 %	3.63 %	2.07 %
01020	Aetna Life and Casualty	AK	2,382	16.92	43.60	0.12	0.00	14.77	11.48	23.33	35.57
00825	North Dakota BC/BS	WY	2,315	16.80	36.62	10.44	4.11	19.50	5.06	2.78	1.59
00710	Michigan BC/BS	MI	66,214	15.25	36.98	3.13	0.30	15.73	8.57	6.53	26.38
14330	Group Health Insurance, Inc.	NY	9,609	15.17	32.83	1.51	0.15	17.27	1.91	4.44	3.77
10250	Travelers Insurance Co.	MS	11,209	15.02	24.21	3.10	0.54	9.95	20.00	79.04	85.11
10490	Travelers Insurance Co.	VA	22,058	12.87	21.24	2.78	0.31	11.75	10.00	19.49	24.59
00751	Montana BC/BS	MT	3,739	12.70	22.17	3.16	0.18	14.63	2.00	16.67	1.09
00074	S.S.S.	VI	8	12.50	16.67	0.00	0.00	12.50	0.00	0.00	0.00
01120	Aetna Life and Casualty	HI	5,982	11.40	24.92	0.39	0.00	12.52	8.14	0.00	0.85
16510	Nationwide Insurance	WV	15,163	11.20	17.93	1.64	0.66	12.43	5.81	0.40	2.46
00528	Arkansas BC/BS	LA	30,248	10.53	18.24	1.18	0.72	10.46	3.51	15.28	10.64
00650	Kansas BS	KS	9,445	10.16	24.43	0.18	0.39	5.45	4.70	15.55	37.60
00860	Pennsylvania BS	NJ	37,087	8.74	15.96	0.93	0.83	10.06	5.53	3.17	2.27
10240	Travelers Insurance Co.	MN	15,323	8.72	14.41	2.98	0.45	4.28	20.69	33.42	38.84
00590	Florida BC/BS	FL	106,031	7.93	12.33	2.55	0.96	9.05	6.42	1.00	1.25
00655	Kansas BS	NE	7,045	7.78	21.94	1.63	0.33	5.11	3.61	11.40	27.90
01380	Aetna Life and Casualty	OR	17,800	6.71	11.95	0.54	0.22	3.05	8.13	19.72	32.79
00550	Colorado BC/BS	CO	16,308	6.52	18.01	0.98	0.25	7.32	4.82	1.98	1.15
00870	Rhode Island Blue Shield	RI	5,925	5.00	7.68	2.79	1.37	4.33	8.62	6.85	8.48
11260	Missouri General Am. Life	MO	23,668	3.37	6.92	1.40	0.62	2.47	2.48	5.57	11.50
00510	Alabama BC/BS	AL	28,724	3.35	5.78	0.16	0.02	3.52	3.36	1.86	0.77
01290	Aetna Life and Casualty	NV	6,008	3.33	8.38	0.44	0.08	3.37	2.59	1.45	3.82
00660	Kentucky BC/BS	KY	21,296	2.96	4.34	0.74	0.09	1.62	3.04	36.81	13.78
00740	Kansas BS	MO-KS	11,636	2.31	4.08	0.66	0.10	1.58	1.11	2.19	8.98
00640	Iowa BS	IA	17,589	2.14	3.61	0.32	0.00	1.93	0.86	2.98	4.28
05440	Connecticut General Life	TN	26,944	1.87	2.66	0.41	0.37	1.83	0.62	1.14	3.12
00542	California BS	No. CA	162,068	1.38	1.98	1.00	0.23	1.27	7.55	2.34	1.40
00973	S.S.S.	PR	10,987	1.21	1.74	0.63	0.00	1.21	0.00	0.50	1.82

SOURCE: Physician Registry, July 1993.

percent to 3.6 percent, the four carriers with the lowest rates in 1989 are same as the four overall lowest carriers.

In 1990-91, the Georgia RRB carrier had the highest MEDU rates with 35.9 percent while the Utah RRB carrier had the third highest at 22.5 percent. The 35.9 percent value for the Georgia operations would have, *ceteris paribus*, placed it as the 30th highest carrier during 1989. Besides the two RRB carriers, Wisconsin Physician Service, TOLIC, and North Dakota BC/BS (North & South Dakota) had MEDU rates greater than the national average of 11.77 percent. The four lowest MEDU rates were less than 0.19 percent; the carriers were: S.S.S. (Virgin Islands), Aetna (Alaska), Alabama BC/BS, and Kansas BS (Kansas).

In 1992-93, the four carriers that had the highest MEDU rates were the same as for 1990-91. The highest rates in 1992-93 were about 10-14 percentage points lower than in 1990-91. There were no other carriers with MEDU rates exceeding the national mean of 6.31 percent. There were six carriers that had MEDU rates equal to zero.

Three of the four carriers that had the highest overall MEDU rates were also highest for medical doctors with Arkansas BC/BS (Arkansas) replacing TOLIC. The four highest MEDU rates ranged from 66.2 percent to 45.6 percent. The carriers with the lowest MEDU rates were S.S.S. (Puerto Rico), California BS, Kansas BS (Missouri-Kansas), and Kentucky BC/BS. None were less than one percent.

For doctors of osteopathy, the Utah RRB, at 65.5 percent, had the highest MEDU rate followed by Connecticut General Life (North Carolina) at 54.6 percent, Wisconsin Physician Service at 53.9 percent, and the Georgia RRB at 53.7 percent. There were four carriers that had MEDU rates equal to zero for osteopaths: S.S.S. (Virgin Islands and Puerto Rico), Maryland BC/BS, and South Carolina BC/BS.

Even though the average MEDU rate for dentists was second lowest among provider types, the highest MEDU rates for dentists were higher than the highest for any other provider type with Wisconsin Physician Service the highest at 95.4 percent. The next three highest rates

for dentists ranged from 82 percent to 79 percent: Minnesota BC/BS and Travelers (Mississippi and Connecticut). Three carriers had MEDU rates equal to zero for dentists: S.S.S. (Virgin Islands), Connecticut General Life (Idaho), and Aetna (Hawaii). Nationwide (West Virginia) at 0.4 percent was the only one other carrier that had a MEDU rate less than 0.5 percent. Although not quite as high as for dentists, the four highest MEDU rates for other limited license practitioners were very high with values ranging from 85.1 percent to 61.8 percent. The four carriers with highest rates were Travelers (Mississippi), Aetna (Oklahoma), Wisconsin Physician Service, and TOLIC. Only S.S.S. (Virgin Islands), had a zero MEDU rate for other limited licensed practitioners. The next three lowest were Connecticut General Life (Idaho) at 0.4 percent, Alabama BC/BS at 0.77 percent, and Aetna (Hawaii) at 0.85 percent.

4.2.4 State License Number

Despite not being a mandatory data element during the implementation phase, state license number had a lower combined MEDU rate than birth year. The integrity requirements for state license number, however, were less strict than for other data elements. The only test for state license number was whether or not it was missing. C & S Administrative Services (Vermont, Massachusetts, and Maine), Travelers (Connecticut) had the highest overall missing rates, ranging from 65.8 percent to 52.8 percent (Table 4-6). There were ten carriers that had overall missing rates equal to zero and another four at 0.01 percent.

Aside from the Georgia RRB carrier, the carriers with highest missing rates during the 1989 implementation year were C & S Administrative Services (Vermont and Maine), Travelers (Connecticut), and Missouri General. The four highest rates ranged from 90.2 percent down to 75 percent. There were 15 carriers that had missing rates equal to zero during 1989.

In 1990-91, the highest missing rates were for the Georgia RRB, Missouri General, Arkansas BC/BS (Arkansas), and the Utah RRB. There were ten carriers than had missing

TABLE 4-6

STATE LICENSE NUMBER, MISSING VALUES (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Name	State	Number of Practice Settings	Overall Rate	*Year Setting Added* 1989	1990-1	1992-3	*Provider Type* MD	DO	Dentist	Other LLP
	National		1,999,067	13.53 %	18.44 %	11.46 %	5.84 %	13.73 %	8.38 %	13.87 %	13.41 %
00781	C&S Admin. Services	VT	4,585	65.78	90.19	18.71	18.02	63.17	57.95	96.30	73.49
10230	Travelers Ins. Co.	CT	17,682	54.74	87.52	23.35	7.39	52.19	50.00	81.65	68.45
00700	C&S Admin. Services	MA	62,426	54.10	70.02	22.29	17.19	53.94	28.46	52.22	62.72
21200	C&S Admin. Services	ME	6,631	52.80	75.00	15.49	10.51	53.18	40.67	82.46	46.56
00520	Arkansas BC/BS	AR	15,821	50.29	74.51	28.01	0.37	54.55	21.74	29.06	11.18
11260	Missouri Gen. Amer. Life	MO	23,668	46.58	84.73	29.63	7.84	45.37	33.28	73.25	66.24
01040	Aetna Life & Casualty	GA	27,908	46.13	71.81	14.65	0.42	49.58	12.27	22.83	13.85
01380	Aetna Life & Casualty	OR	17,800	38.78	62.88	16.03	0.43	38.53	21.19	51.90	45.07
00780	C&S Admin. Services	NH	6,645	38.42	67.14	12.76	14.03	32.29	27.69	89.66	67.10
01370	Aetna Life & Casualty	OK	13,152	34.30	62.68	5.32	0.07	32.25	19.46	61.31	64.61
01360	Aetna Life & Casualty	NM	7,337	32.98	70.43	6.46	0.06	30.75	23.17	67.77	53.28
10074	Travelers-RRB-UT	RRB	143,270	30.36	72.08	23.88	20.37	30.69	65.15	27.73	26.19
00690	Maryland BS	MD	25,740	29.97	56.10	0.53	0.21	32.60	0.00	0.25	0.27
01030	Aetna Life & Casualty	AZ	20,462	27.77	53.42	13.44	0.65	24.78	34.55	7.18	47.34
10072	Travelers-RRB-GA	RRB	183,394	26.55	100.00	36.60	11.02	26.30	53.40	23.77	28.05
01020	Aetna Life & Casualty	AK	2,382	26.41	56.40	13.36	0.00	25.24	13.11	46.67	39.53
00570	Pennsylvania BS	DE	3,557	25.22	38.81	3.11	0.94	28.27	15.88	18.50	6.18
01120	Aetna Life & Casualty	HI	5,982	23.91	52.54	0.48	0.00	26.16	16.28	0.00	2.99
00580	Pennsylvania BS	DC	20,151	23.80	36.98	5.06	0.68	24.26	14.91	36.91	10.45
10250	Travelers Ins. Co.	MS	11,209	21.87	36.51	2.00	0.11	17.37	30.00	84.72	81.25
10240	Travelers Ins. Co.	MN	15,323	21.65	38.64	1.89	0.26	18.48	41.38	43.88	41.92
00865	Pennsylvania BS	PA	83,361	19.01	29.34	5.17	0.72	21.34	18.36	13.71	4.66
10490	Travelers Ins. Co.	VA	22,058	18.75	32.33	1.20	0.18	17.69	20.00	26.47	29.30
01290	Aetna Life & Casualty	NV	6,008	18.34	46.03	2.66	0.16	17.37	3.54	34.78	37.48
16510	Nationwide Ins. Co.	WV	15,163	18.12	29.60	1.45	0.43	18.61	4.24	0.40	31.70
00801	BS of Western NY	NY	28,702	15.95	29.97	0.42	0.10	17.85	17.87	0.15	0.04
00528	Arkansas BC/BS	LA	30,248	10.04	17.48	1.14	0.49	9.37	3.51	32.81	16.06
00803	Empire BC/BS	NY	92,461	9.69	12.50	0.10	0.66	10.90	2.55	0.00	1.00
00650	Kansas BS	KS	9,445	9.18	22.18	0.27	0.00	9.76	4.07	9.64	8.99

TABLE 4-6 (continued)

STATE LICENSE NUMBER, MISSING VALUES (PERCENTAGE) RATES, BY CARRIER IN DESCENDING ORDER OF THE OVERALL RATE

Carrier Number	Name	State	Number of Practice Settings	Overall Rate	Year Setting Added			Provider Type			
					1989	1990-1	1992-3	MD	DO	Dentist	Other LLP
00655	Kansas BS	NE	7,045	9.08 %	24.94 %	2.54 %	0.00 %	9.17 %	2.41 %	8.77 %	9.26 %
00860	Pennsylvania BS	NJ	37,087	8.59	16.07	0.64	0.08	9.44	8.54	2.45	3.69
16360	Nationwide Ins. Co.	OH	84,069	7.13	12.83	0.10	0.05	7.06	2.35	16.01	11.01
00910	Utah BC/BS	UT	6,432	6.37	11.25	0.89	0.09	3.42	0.00	15.34	29.86
00751	Montana BC/BS	MT	3,739	6.15	11.67	0.08	0.00	7.55	0.00	0.00	0.00
00870	Rhode Island BS	RI	5,925	5.22	9.08	1.50	0.58	2.10	3.45	16.98	19.35
00740	Kansas BS	MO	11,636	4.30	2.64	9.46	0.05	4.01	6.02	7.10	4.49
05440	Connecticut General	TN	26,944	0.99	0.52	2.88	0.37	0.60	0.31	0.71	7.56
14330	Group Health Ins., Inc.	NY	9,609	0.31	0.26	0.56	0.00	0.34	0.00	0.00	0.18
05535	Connecticut General	NC	40,046	0.17	0.14	0.42	0.00	0.19	0.00	0.00	0.04
00621	Illinois BC/BS	IL	70,990	0.08	0.05	0.15	0.05	0.08	0.00	0.00	0.06
00951	Wisconsin Physician Serv.	WI	24,456	0.08	0.13	0.02	0.02	0.10	0.00	0.00	0.03
00720	Minnesota BC/BS	MN	14,907	0.07	0.03	0.14	0.04	0.06	0.00	0.00	0.18
00590	Florida BC/BS	FL	106,031	0.05	0.01	0.11	0.10	0.02	0.06	0.11	0.27
00620	North Dakota BC/BS	SD-ND	7,501	0.03	0.00	0.04	0.09	0.02	0.00	0.45	0.00
00550	Colorado BC/BS	CO	16,308	0.02	0.00	0.06	0.00	0.03	0.00	0.00	0.00
00640	Iowa BS	IA	17,589	0.01	0.00	0.04	0.00	0.01	0.00	0.00	0.05
00880	South Carolina BC/BS	SC	24,095	0.01	0.00	0.05	0.00	0.01	0.00	0.00	0.00
00932	Washington Physician Serv.	WA	23,763	0.01	0.03	0.00	0.00	0.02	0.00	0.00	0.00
00973	S.S.S.	PR	10,987	0.01	0.00	0.00	0.07	0.01	0.00	0.00	0.00
00510	Alabama BC/BS	AL	28,724	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00542	California BS	No. CA	162,068	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00630	Indiana BC/BS	IN	28,863	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.04
00660	Kentucky BC/BS	KY	21,296	0.00	0.00	0.03	0.00	0.00	0.00	0.27	0.00
00710	Michigan BC/BS	MI	66,214	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
00825	North Dakota BC/BS	WY	2,315	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00900	Texas BC/BS	TX	96,351	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
00974	S.S.S.	VI	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
02050	Transamerica Occidental	So. CA	117,191	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
05130	Connecticut General	ID	4,307	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

SOURCE: Physician Registry, July 1993.

rates equal to zero in 1990-91. In 1992-3, the Utah RRB had the highest missing rate followed by three of C & S Administrative Services' states: Vermont, Massachusetts, and New Hampshire. There were 20 carriers that had missing rates equal to zero during 1992-93.

Except for Arkansas BC/BS (Arkansas) replacing Travelers (Connecticut), the carriers with the four highest missing rates for medical doctors were the same as for the overall rate. The missing value rates for these four carriers ranged from 63.2 percent to 53.2 percent. There were ten carriers that had missing value rates equal to zero for medical doctors. For osteopaths, the carriers with the highest rates were the two RRB carriers, C & S Administrative Services (Vermont), and Travelers (Connecticut). There were 24 carriers that had missing values rates equal to zero for osteopaths.

The four highest missing value rates for dentists are substantially higher than for medical doctors and osteopaths with values ranging from 96.3 percent to 82.5 percent. C & S Administrative Services (Vermont, New Hampshire, and Maine) and Travelers (Mississippi) had the highest rates. There were 21 carriers that had missing values rates equal to zero for dentists. Although not quite as high as the highest dentist rates, the missing value rates for other limited license practitioners were also much higher than for medical doctors and osteopaths. Once again, C & S Administrative Services (Vermont and New Hampshire) and Travelers (Mississippi and Connecticut) had the highest rates. There were 15 carriers that had missing values rates equal to zero for other limited license practitioners.

4.2.5 Summary

There were a number of instances in which the *relative* performance of individual carriers dramatically improved or decreased over time. For instance, in 1989 Aetna (Georgia) had the highest MEDU rate for school codes; by 1992-93, its MEDU rate was close to the median. For birth year, Empire BC/BS was in the top third of carriers with the highest MEDU

rate in 1989, by 1992-93, they were in the bottom third. Conversely, despite only a slight increase in its MEDU rate, California BS went from the second lowest birth year MEDU rate to among the top third.

There were also instances in which the relative rankings did not change very much. For graduation year, Wisconsin had the highest MEDU rate in 1989[5] and the second highest rates in 1990-91 and 1992-93. For birth year, Aetna (Georgia) had the highest MEDU rate in 1989, sixth highest in 1990-91, and 12th highest in 1992-93. The identities of some of the carriers with the very highest and very lowest MEDU rates often changed over time. That is, while two of the carriers among the four highest (four lowest) usually remained among the four highest (four lowest), the identities of the other carriers changed over time.

Despite some changes in rankings, carriers with higher than average MEDU rates in 1989 usually had higher than average MEDU rates in 1992-3. The correlation coefficients ranged from 0.38 for school code to 0.63 for state license number between 1989 and 1992-3. Note, however, that by 1992-93, the range of MEDU rates was significantly smaller than in 1989. For instance, the MEDU rate in 1989 for birth year had a minimum value of zero percent and a maximum value of 89 percent.[6] By 1992-93, the range was 0-19 percent, with two carriers having MEDU rates higher than two percent, three carriers between one and two percent, and over two-thirds of the carriers at less than 0.5 percent.

We found, using correlation analysis, that carriers that had higher than average MEDU rates for one provider type usually had higher than average rates for the other provider types. For graduation year, for instance, the correlation coefficients of MEDU rates for medical doctors was 0.46 with osteopaths, 0.48 with dentists, and 0.35 with other limited license practitioners. Finally, carriers that had higher than average MEDU rates for one data element usually also had higher than average rates for the other three data elements. For instance, the correlation

[5]The Georgia RRB actually had a higher rate in 1989 but only submitted two practice settings to the Registry.

[6]Ignoring, again, the 100 percent rate of the Georgia RRB.

coefficients of the overall MEDU rates for birth year was 0.83 with school code, 0.66 with graduation year, and 0.51 with state license number.

Two factors seem primarily associated with high overall MEDU rates for the four data elements: (1) government restrictions were placed on carriers in obtaining information directly from providers or from copyrighted sources during the 1989 implementation phase of the Registry and (2) school code, graduation year, and state license number did not become mandatory data elements until December 1991. Prior to the establishment of the Registry, many carriers did not collect the above data elements because they were not needed for payment purposes. Carriers, thus, often were not able to obtain the desired information during 1989. Removal of restrictions in obtaining information directly from providers and changes in making data elements mandatory helped significantly lower MEDU rates by 1992-93. Increased carrier experience in obtaining information and submitting records to the Registry also probably helped to lower MEDU rates by 1992-93.

Despite improvements in MEDU rates, it nonetheless remains the fact that practice settings added to the Registry in 1989 still have high MEDU rates. As shown in Tables 4-3 through 4-6, many carriers have not corrected or updated the values of data elements for settings added in 1989. During our site visits to carriers in the summer of 1993, several carriers indicated that it was difficult to obtain accurate information from providers once a UPIN had been issued. They stated that they only had leverage on the providers to obtain such information when a provider initiated a change to an existing practice setting or wanted to add a new practice setting. The high MEDU rates for 1989, however, also suggest that carriers might not have placed a high priority in retroactively obtaining the requisite information. Another possibility is that carriers have not always submitted update records to the Registry for information they currently have in their databases but did not have during 1989. We note that while Transmittal No. 6 (December 1991) for the *Medicare Carriers Manual* (Part 4 - Professional Relations) changed the three data elements from "if available" to required, it did

not explicitly instruct carriers to obtain or submit such information for practice settings that had already been incorporated into the Registry.

While MEDU rates have significantly declined since 1989, the MEDU rates for the two RRB carriers remain conspicuously higher than for other carriers. As discussed in Chapter 2, the RRB carriers do not have the same access as other carriers to high quality provider information. While the RRB carriers might be able to lower their MEDU rates still more, it seems likely that they will continue to have the highest rates.

4.3 Carrier-Specific Analysis

In this section we describe the MEDU rates of individual carriers and contrast them to the national averages. Almost all carriers had high MEDU rates, in both the header and practice settings versions, for birth year, medical school codes, and graduation year. In most, but certainly not all carriers, MEDU rates decline from high initial levels associated with the 1989 implementation phase to much lower levels in 1992-3. While the decline in MEDU rates over time is encouraging, it needs to be emphasized that there are many errors still present in the Registry data. Tables for the individual carriers are provided in Appendix A.

Prior to the analysis of the carrier data, we somewhat expected that the highest MEDU rates would be isolated among a few carriers. The results indicate that the problems are pervasive and are not limited to just a few carriers. Additionally, over the course of time, the carriers with the highest MEDU rates for a specific variable have changed.

Caution is required in interpreting MEDU rates for the header data elements for each individual carrier. The reason is due to cases in which multiple carriers service a provider. In such instances, any one of the carriers could have supplied the value of a header data element. It, therefore, might be inappropriate to identify a MEDU header value with a specific carrier. It is likely, however, that most header values were taken from the first practice setting added to

the Registry. Since the carrier code was taken from the first practice setting, the MEDU rates reported for the header data elements in the tables for each carrier are probably indicative of those MEDU values for which the carrier is responsible. The MEDU rates reported for the setting versions of data elements are not subject to ambiguity.

Each carrier vignette has the same order of discussion. The number of data elements whose combined MEDU rate for the "overall" column exceeds 0.5 percent is presented first.[7] The names of these data elements are then listed. Next, for each individual MEDU component (e.g., missing value), data elements whose rates exceed 0.5 percent are listed. Combined MEDU rates by DRIP status are discussed if the de-activated value causes the overall rate to exceed 0.5 percent or if the de-activated rate is two or more times as high as the active rate. Then time trend of the combined MEDU rates for the data elements is described. Combined MEDU rates by provider type are characterized next. Finally, the carrier's overall MEDU rates are compared to the national averages.

Carrier 510: Alabama BC/BS

Most of the rates indicating poor variable integrity are low (Table A-1). Six data elements, however, have combined MEDU rates 0.5 percent and above.[8] The data elements (in both their header and practice setting versions) are birth year, medical school, and graduation year. Missing value rates are high for all six. The dubious value rate is equal to 0.95 percent for the header version of graduation year.

Most of the MEDU values are associated with records added in 1989 (except header medical school which is poor for settings added in 1990-91). MEDU rates are high for MDs and DOs for all six of these variables and, for graduation year, there are also problems with data for dentists and other LLPs.

[7] Any MEDU rate exceeding the 0.5 percent standard is considered "high" for the purposes of these vignettes.

[8] The 0.5 percent value is referred to, in the text, as the (discussion) *standard*.

The MEDU rates of the six variables for this carrier are much better than the national averages. Additionally, in contrast to the national MEDU rates for presence of a license and the indicator for practice type, Alabama BC/BS has negligible MEDU rates.

Carrier 520: Arkansas BC/BS (Arkansas)

Most of the rates indicating poor variable integrity are low (Table A-2). Seven data elements (variables), however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is state license number.

Except for error rates for medical school codes, missing values are the principal MEDU value. For instance, the missing value rate for state license number is 50 percent. Most MEDU values are associated with records added in 1989, with declines over time in nearly all MEDU rates except for error rates for medical school codes. Note, however, that the incidence of missing values for medical school codes declines to near zero in 1992-3. While the MEDU rates are high for all of the above variables by credential, the MDs have, by far, the highest MEDU rates. Arkansas BC/BS's MEDU rates for the seven variables are considerably higher than the national average.

Carrier 528: Arkansas BC/BS (Louisiana)

Most of the rates indicating poor variable integrity are low (Table A-3). Eight variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other two are state license number and the header version of death year.

Missing values and dubious values are the most frequent MEDU types. Missing value rates well exceed 0.5 percent for all eight variables except for death year. Most of the MEDU rates decline over time, especially missing values and dubious values. The decline in medical

school missing value rates is partially offset by a small increase in the error rates. Missing values are high for state license numbers; however, the rate declines over time. For each of the nine variables, missing value rates are higher than the standard for each of the provider types.

In most instances, the carrier MEDU rates are lower than the national average. Since Arkansas BC/BS is the carrier for both Arkansas and Louisiana, it is somewhat puzzling that the MEDU rates in Louisiana are quite a bit lower than in Arkansas.

Carrier 542: California BS (Northern California)

Most of the rates indicating poor variable integrity are low (Table A-4). This carrier has an unusually high number (336) of practice settings with a missing DRIP status (but only .21 percent of all practice settings). At the *overall* level, only graduation year (in both header and practice setting versions) have MEDU rates exceeding the standard. These are for dubious and missing values which tend to decline over time. Each of the credential groups had high dubious and/or missing value rates for graduation year. Dubious values for death rate exceeds 0.5 percent only in 1989. Finally, there seems to be difficulties in obtaining accurate medical school information for DOs. California Blue Shield's overall problem rates are generally lower than the national averages.

Carrier 550: Colorado BC/BS

Most of the rates indicating poor variable integrity are low (Table A-5). Six variables, however, have combined MEDU rates exceeding the standard. The variables (in both their header and practice setting versions) are birth year, medical school, and graduation year with high missing value rates for each of the six. The missing value rates decline over time from the high initial problem rates in 1989. MDs and DOs have the highest MEDU rates for the six variables. Colorado BC/BS's MEDU rates are generally lower than the national averages.

Carrier 570: Pennsylvania BC/BS (Delaware)

Most of the rates indicating poor variable integrity are low (Table A-6). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is state license number. Missing value rates are high for all seven variables and dubious value rates are high for both versions of graduation year.

The MEDU rates for de-activated records are substantially higher than for active records. The percentage difference is greater for the header versions of the variables than for the setting versions. This implies that the problems are concentrated in providers who have more practice settings. The MEDU rates decline from very high levels in 1989 to rates below 0.5 percent (except for state license number). However, the number of practice settings (427) added during 1992-3 is low. The missing value rates for MDs and DOs are high relative to other providers. The rate of unknown primary specialty codes is very high (3.47 percent) for dentists.

Compared to the national averages, the MEDU rates for the header versions of birth year, medical school, and graduation year are lower. The MEDU rates for the four settings variables, however, are generally higher than the national averages.

Carrier 580: Pennsylvania BC/BS (District of Columbia)

Most of the rates indicating poor variable integrity are low (Table A-7). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other is state license number. Missing value rates are high for all variables, dubious value rates are high for both versions of graduation year, and erroneous value rates are high for both versions of medical school.

The pattern of MEDU rates by DRIP status are the same for the District of Columbia as for Delaware (Pennsylvania BC/BS is the carrier for both). MEDU rates usually decline from very high levels in 1989 to rates lower than the standard in 1992-3; exceptions are the two medical school variables in which the error rates increase (as the missing rates decline). Although the MEDU rates exceed the standard for all provider types, they are usually the highest for MDs and DOs. An exception is primary specialty in which dentists are the only group that have rates exceeding the standard for unknown values and erroneous values.

The MEDU rates for the District of Columbia are lower than the national averages for the header variables. For the practice setting variables, the District of Columbia MEDU rates are fairly close to the national averages with several rates that are substantially higher and several rates that are substantially lower than the national averages.

Carrier 590: Florida BC/BS

Most of the rates indicating poor variable integrity are low (Table A-8). Six variables, however, have combined MEDU rates 0.5 percent and above. The variables (in both their header and practice setting versions) are birth year, medical school, and graduation year. Missing value rates are high for all variables, dubious value rates are high for both versions of graduation year, and erroneous value rates are high for the header version of medical school.

MEDU rates are significantly higher for de-activated settings than for active settings. Although MEDU rates decline from the high levels of 1989, they still exceed 0.5 percent for medical school (header) and graduation year (missing rates for both header and setting). MEDU rates exceed the standard for each of the provider types for at least one of the six variables. MDs and DOs usually have the highest MEDU rates. For both header and setting variables, Florida's MEDU rates are usually lower than the national average.

Carrier 621: Illinois BC/BS

Most of the rates indicating poor variable integrity are low (Table A-9). Six variables, however, have combined erroneous and missing rates 0.5 percent and above. The variables (in both their header and practice setting versions) are birth year, medical school, and graduation year. Missing value rates are relatively high for all six variables and erroneous values are relatively numerous for both versions of medical school.

MEDU rates for de-activated settings are somewhat higher than for active settings. MEDU rates decline over time from fairly high levels in 1989 to less than 0.5 percent in 1992-3 (except for medical school error rates). Missing values for the six variables are very high for all provider types for graduation year (except DOs which also have a value exceeding the standard). Missing value rates are also high for MDs for birth year and medical school (both header and setting versions).

For the six variables discussed thus far, Illinois' combined MEDU rates exceed the national averages. At the national level, MEDU rates for state license number and group practice indicator exceed 0.5 percent, but not in Illinois.

Carrier 630: Indiana BC/BS

Most of the rates indicating poor variable integrity are low (Table A-10). Six variables, however, have combined MEDU rates 0.5 percent and above. The six variables (in both their header and practice setting versions) are birth year, medical school, and graduation year. Missing value rates are high for *the* six variables,[9] dubious value rates are high for both versions of graduation year, and erroneous value rates are high for the header version of medical school.

The MEDU rates decline from high levels in 1989 to less than the standard for all except the header version of medical school. The MEDU rates exceed the standard for all provider

[9]When "the" is italicized and precedes "six," the variables referred to are birth year, medical school codes, and graduation year.

types for at least one variable with very high rates for MDs and other LLPs. For the header versions of variables, Indiana's MEDU rates are usually lower than the national averages. For the setting versions, however, Indiana usually has higher rates.

Carrier 640: Iowa BS

Most of the rates indicating poor variable integrity are low (Table A-11). Five variables, however, have combined MEDU rates 0.5 percent and above. Four of the variables (in both their header and practice setting versions) are medical school and graduation year; the fifth is the setting version of birth year. Missing values are high for all five variables, dubious values are high for both versions of graduation year, and erroneous values exceeded one percent for both versions of medical school.

The MEDU rates decline over time leaving only two rates above the (discussion) standard. Missing values and erroneous values for medical school are the highest for MDs and DOs. Graduation year MEDU rates exceeded 0.5 percent for all provider types. Except for medical school codes (both header and setting versions), Iowa's MEDU rates are lower than the national averages.

Carrier 650: Kansas BS (Kansas)

Most of the rates indicating poor variable integrity are low (Table A-12). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is state license number. Missing values are well above one percent for *the* six variables and the state license number. The dubious value rate equals 0.5 percent for graduation year (header version). Erroneous values exceed five percent for both versions of medical school.

MEDU rates usually decline over time from the high rates in 1989. Medical school and graduation year MEDU rates are high for all provider types with dentists and other LLPs having exceptionally high rates. Missing value rates exceed three percent for MDs and osteopaths for several variables. Missing values are also quite high for all types of providers. Compared to the national averages, Kansas' *overall* MEDU rates are often lower. The exceptions are usually by provider type.

Carrier 655: Kansas BS (Nebraska)

Most of the rates indicating poor variable integrity are low (Table A-13). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is state license number. All seven variables have high missing value rates.

MEDU rates usually decline from high rates in 1989. By 1992-3, only error rates for medical school (both header and setting) exceed 0.5 percent. The rise in medical school (setting) error rates might be connected to the decline of the missing value rate to zero. All seven variables have high MEDU rates for at least one of the provider types with other LLPs usually having the highest rates. Compared to the national averages, MEDU rates for Nebraska are usually lower.

Carrier 660: Kentucky BC/BS

Most of the rates indicating poor variable integrity are low (Table A-14). Six variables, however, have combined MEDU rates 0.5 percent and above. The six variables (in both their header and practice setting versions) are birth year, medical school, and graduation year. Missing values are high for all except birth year. Erroneous values are high for both versions of birth year and medical school.

Except for medical school (setting) error rates, MEDU decline over time. After a large decline from 1989 to 1990-1, medical school error rates nearly doubles to 1.71 percent. There are MEDU rates for each of the provider types for a least one variable. For medical school (both), MDs and DOs have the highest error rates while dentists and other LLPs have the highest missing value rates. For DOs the erroneous rate exceeds two percent. As discussed in Chapter 3, Kentucky has 1,170 missing DRIP values (nearly 61 percent of the national cases) which represents 5.49 percent of Kentucky's total practice settings.

With exception of DRIP status, for *overall* MEDU rates Kentucky has lower values than the national averages. For some specific categories (e.g., medical school missing values for dentists), however, Kentucky has MEDU rates much higher than the national averages.

Carrier 690: Maryland BS

Most of the rates indicating poor variable integrity are low (Table A-15). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other is state license number. All seven variables have high missing value rates.

The high initial MEDU rates each declined to below 0.5 percent by 1992-3. Missing value rates for birth year, medical school, and graduation year are highest for MDs by a substantial margin. For DOs, the rate of erroneous values exceed 0.5 percent for primary specialty. Maryland's *overall* MEDU rates are similar to the national averages.

Carrier 700: C&S Administrative Services (Massachusetts)

Most of the rates indicating poor variable integrity are low (Table A-16). Eight variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other two are primary specialty code and state license number. Missing value rates are high

for all eight except primary specialty code. Erroneous value rates are high for primary specialty code. Dubious value rates are high for the header version of graduation year.

Combined MEDU rates decline over time except for primary specialty code. Missing value rates for state license number are higher than 25 percent for all provider types. Aside from state license number, each provider type has at least one combined MEDU rate that exceeds two percent for at least one variable. MEDU rates for Massachusetts are substantially higher than the national averages.

Carrier 710: Michigan BC/BS

Most of the rates indicating poor variable integrity are low (Table A-17). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is group practice indicator. Missing values are high for *the* six. Erroneous values are high for medical school (both) and group practice indicator.

Except for medical school codes (both), there is a decline in the MEDU rates over time. For medical schools, the missing values rates continuously declined while the error rates peaked in 1990-1 and then declined, albeit at levels higher than in 1989. For graduation year, missing values are high for all provider types with other LLPs especially high. For medical school codes, MDs have the highest missing rate while DOs have the highest error rate. Michigan usually has MEDU rates lower than the national averages.

Carrier 720: Minnesota BC/BS (Rural Minnesota)

Most of the rates indicating poor variable integrity are low (Table A-18). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation

year; the seventh is group practice indicator. For *the* six variables, missing value rates are high. Erroneous value rates are quite high for group practice indicator.

MEDU rates (especially missing values) were especially high in 1989 but declined to below the (discussion) standard by 1992-3. MDs and DOs have especially high missing rates for birth year (both header and setting), medical school (both), and graduation year (both). Indeed, graduation year has very high missing value rates for all provider types. Similarly, the group practice indicator has very high error rates across all provider types. On an *overall* basis, compared to the national averages, Minnesota BC/BS usually has higher MEDU rates.

Carrier 740: Kansas BS (Missouri-Kansas)

Most of the rates indicating poor variable integrity are low (Table A-19). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the seventh is state license number. This carrier also has an unusually high number (86) of practice settings with a missing DRIP status.

Missing values are high for all seven variables with especially high rates when DRIP is missing for both versions of medical school and graduation year. Erroneous values are high for medical school codes (both versions). Dubious values are high for graduation year (both). Except for both medical school code variables, problem rates decline over time. For both medical school code variables, missing value rates decline from high values in 1989 to less than 0.5 percent. However, as medical school missing rates fell, the error rates peaked in 1990-1 and then declined, albeit at a higher level than in 1989.

Each provider type had at least one high MEDU rate with each also having the highest value for at least one of the seven variables. With the major exception of DRIP status, Carrier 740 generally has MEDU rates lower than the national averages.

Carrier 751: Montana BC/BS

Most of the rates indicating poor variable integrity are low (Table A-20). Eight variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other two are primary specialty and state license number. Missing value rates are high for all eight except primary specialty. Erroneous values are high for medical school code (both versions) and primary specialty. Both versions of graduation year have dubious values exceeding 0.5 percent.

MEDU rates usually decline over time. For both medical school code variables, missing value rates decline from high values in 1989 to zero. However, as medical school missing rates fell, the error rates peaked in 1990-1 and then declined with the header error rate higher in 1992-3 than in 1989 and setting error rate lower in 1992-3 than in 1989. Each provider type has at least one variable in which a MEDU rate exceeds 0.5 percent – often by a substantial margin. Errors in primary specialty codes are a particular problem for MDs and DOs. Compared to the national averages, Montana usually has lower MEDU rates.

Carrier 780: C & S Administrative Services (New Hampshire)

Most of the rates indicating poor variable integrity are low (Table A-21). Eight variables, however, have missing rates 0.5 percent and above. The variables are birth year, medical school, and graduation year in both their header and practice setting versions as well as state license number and primary specialty. Missing variables are high for all eight except primary specialty.

MEDU rates for de-activated settings are substantially higher than for active settings. However, the MEDU rates of active settings substantially exceed four percent except for primary specialty. As is usually the case, most MEDU rates fell dramatically over time. Primary specialty has some problems with erroneous data among MDs, although rates are

under a 1 percent level. State license number has high missing value rates for all provider groups. Overall, variable integrity rates are superior to national averages in the overall category except for two variables -- primary specialty and state license number.

Carrier 781: C & S Administrative Services (Vermont)

Most of the rates indicating poor variable integrity are low (Table A-22). Eight variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other two are primary specialty code and state license number. Missing value rates are high for all eight except primary specialty code. Erroneous value rates are high for primary specialty code. Dubious value rates are high for both versions of graduation year.

Combined MEDU rates decline over time. Missing value rates for state license number exceed 50 percent for all provider types. Aside from state license number, each provider type has at least one combined MEDU rate that exceeds 14 percent for at least one variable. MEDU rates (the five practice setting variables only) for Vermont are substantially higher than the national averages.

Carrier 801: BC/BS of Western New York

Most of the rates indicating poor variable integrity are low (Table A-23). Eight variables, however, have combined erroneous and missing rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other two are primary specialty and state license number. Missing values are high for all eight variables except primary specialty. Erroneous values are high for primary specialty. There are high rates of dubious graduation year values.

MEDU rates usually decline over time. Both medical school codes are an exception with error rates increasing as the missing value rates declined. Each provider type has at least

one variable in which a MEDU rate exceeds 0.5 percent – often by a substantial margin. Errors in primary specialty codes are particularly high for DOs and dentists.

At the *overall* level, Carrier 801 has MEDU rates close to the national averages. Carrier 801, however, has higher MEDU rates for primary specialty codes and state license numbers.

Carrier 803: Empire BC/BS (eastern New York)

Most of the rates indicating poor variable integrity are low (Table A-24). Eight variables, however, have combined erroneous and missing rates 0.5 percent and above. Six of the variables (in both their header and practice setting versions) are birth year, medical school, and graduation year; the other two are primary specialty and state license number. Missing values are a problem for all variables except primary specialty graduation year (setting). Erroneous values are high for primary specialty. There are high rates of dubious graduation year values.

MEDU rates usually decline over time. For birth year (both versions) and medical school missing values, the rates reach their nadir in 1990-1 and increase slightly in 1992-3. Each provider type has at least one variable in which a MEDU rate exceeds 0.5 percent – often by a substantial margin. Errors in primary specialty codes are particular noticeable for DOs. Empire's MEDU rates usually exceeded the national averages by a substantial margin.

Carrier 820: North Dakota BC/BS (North and South Dakota)

Most of the rates indicating poor variable integrity are low (Table A-25). Nine variables, however, have MEDU 0.5 percent and above. The variables are birth year, medical school, graduation year, and sanction code in both their header and practice setting versions as well as group practice indicator. Missing values are consistently high for both versions of the birth year, medical school, and graduation year variables. Both versions of medical school code have

error rates exceeding one percent. Unknown sanction code rates are considerably higher than the national average, especially for the setting version.

Combined MEDU rates tended to decline over time except birth year (header) which was higher in 1992-93 than in 1990-91. All four provider categories have high rates of missing medical school codes and missing graduation years, whereas only MDs and DOs have high missing value rates for the birth year and sanction code variables.

Carrier 825: North Dakota BC/BS (Wyoming)

Most of the rates indicating poor variable integrity are low (Table A-26). Four variables, however, have MEDU rates 0.5 percent and above. The variables are medical school and graduation year in both their header and practice setting versions. Missing value rates are high for both medical school and graduation year, though still below national averages. Erroneous value rates, on the other hand, are higher than national averages among medical school and birth year variables. Note that dentists and other LLPs are the major error contributors for the birth year variable. The lack of improvement over time for medical school (header) could be due to another carrier.

Carrier 860: Pennsylvania Blue Shield (New Jersey)

Most of the rates indicating poor variable integrity are low (Table A-27). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Dubious value rates are high for both versions of graduation year.

Combined MEDU rates decline over time. Missing value rates for state license number exceed two percent for all provider types. Aside from state license number, each provider type

has at least one combined MEDU rate that exceeds two percent for at least one variable. MEDU rates for New Jersey are substantially lower than the national averages.

Carrier 865: Pennsylvania Blue Shield (Pennsylvania)

Most of the rates indicating poor variable integrity are low (Table A-28). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Dubious value rates are high for both versions of graduation year.

Combined MEDU rates decline over time. Missing value rates for state license number exceeds four percent for all provider types. Aside from state license number, each provider type has at least one combined MEDU rate that exceeds two percent for at least one variable. The combined MEDU rate for primary specialty code is high for dentists at over five percent. MEDU rates for Pennsylvania tend to be near the national averages.

Carrier 870: Rhode Island BS

Most of the rates indicating poor variable integrity are low (Table A-29). Seven variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven except the header version of birth year. Erroneous value rates are high for both versions of the medical school code. Dubious value rates are high for both versions of graduation year.

Combined MEDU rates declined over time for all variables except the header version of birth year. For state license number, the missing value rate exceeded two percent for each provider type. Additionally, for each provider category, the combined MEDU rate exceeded two percent for at least one other variable. Also, the dubious value rate for graduation year

exceeds two percent for osteopaths. MEDU rates for Rhode Island are substantially lower than the national averages.

Carrier 880: South Carolina BC/BS

Most of the rates indicating poor variable integrity are low (Table A-30). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is group practice indicator. Missing value rates are high for all seven except group practice indicator. Erroneous value rates are high for the header version of the medical school code and group practice indicator. Dubious value rates are high for the header version of graduation year.

Combined MEDU rates decline over time although group practice indicator peaked at a very high rate during 1990-91. For each provider, the combined MEDU rate exceeded ten percent for at least one variable. MEDU rates for South Carolina are substantially higher than the national averages.

Carrier 900: Texas BC/BS

Most of the rates indicating poor variable integrity are low (Table A-31). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is group practice indicator. Missing value rates are high for all seven except group practice indicator. Erroneous value rates are high for the medical school code (both) and group practice indicator. Dubious value rates are high for the header version of graduation year.

Combined MEDU rates decline over time although group practice indicator peaked during 1990-91. Although exceeding 0.5 percent, dentists usually had the lowest combined MEDU rates with 3.4 percent the highest (setting version of graduation year). The other

provider types had a least one combined MEDU rate exceeding 20 percent for at least one variable. The combined MEDU rates for Texas tended to be slightly higher than the national averages.

Carrier 910: Utah BC/BS

Most of the rates indicating poor variable integrity are low (Table A-32). Eight variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other two are state license number and group practice indicator. Missing value rates are high for all eight except group practice indicator. Erroneous value rates are high for birth year (setting) and group practice indicator.

Combined MEDU rates decline over time. For each provider type, the combined MEDU rate exceeds five percent for at least one of the eight variables. Except for state license number, the practice settings versions of birth year, medical school code, and graduation year, the combined MEDU rates for Utah are substantially higher than the national averages.

Carrier 932: Washington Physician Service

Most of the rates indicating poor variable integrity are low (Table A-33). Four variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of medical school code and graduation year. Missing value rates are high for all four.

Combined MEDU rates decline over time. For each provider type, the combined MEDU rate exceeds six percent for at least one of the four variables and are quite high for dentists and other limited license practitioners. Except for the four variables, Washington's combined MEDU rates are lower than the national averages (when the national average is greater than zero).

Carrier 951: Wisconsin Physician Service

Most of the rates indicating poor variable integrity are low (Table A-34). Six variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year. Missing value rates are high for all six except the header version of the medical school code. The missing value rate for graduation year (both) are exceedingly high for all provider types. Erroneous value rates are high for the practice setting version of the medical school code.

Combined MEDU rates considerably decline over time except for the header version of the medical school code. Except for graduation year (both), the MEDU rates for dentists and other limited license practitioners are low. The combined MEDU rates for MDs exceed the 0.5 percent standard for both versions of birth year and the medical school code. Except for the header version of birth year, osteopaths have combined MEDU rates exceeding the standard for all six variables. Except for graduation year, the combined MEDU rates for Wisconsin are lower than the national averages.

Carrier 973: S.S.S. (Puerto Rico)

Most of the rates indicating poor variable integrity are low (Table A-35). Five variables, however, have combined MEDU rates 0.5 percent and above. Four of the variables are (both header and practice settings) of birth year and graduation year; the other is group practice indicator. The erroneous value rate for the group practice indicator is excessively high at nearly 22 percent and are high for all provider types except osteopaths. Dubious value rates are slightly high (between 0.5 and 1.0 percent) for both versions of graduation year.

Combined MEDU rates decline over time except for group practice indicator which reached its highest rate in 1992-93. There are no MEDU values for the two osteopaths. Aside from the group practice indicator and osteopaths, for each of the other provider types, there is

at least one combined MEDU rate exceeding 0.5 for one the four other variables. Except for the group practice indicator, the combined MEDU rates for Puerto Rico are substantially lower than the national averages.

Carrier 974: S.S.S. (Virgin Islands)

Most of the rates indicating poor variable integrity are low (Table A-36). Two variables, however, have overall missing value rates equal to 12.5 percent: both the header and practice settings versions of graduation year. The missing values are only present in the practice settings added in 1989. MDs are the only provider type with missing values. With the exception of both versions of the graduation year, all of the Virgin Islands combined MEDU rates are equal to zero.

Carrier 1020: Aetna Life and Casualty (Alaska)

Most of the rates indicating poor variable integrity are low (Table A-37). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven.

Combined MEDU rates decline over time. For each provider type, the combined MEDU rate exceeds the 0.5 percent standard by a substantial amount for at least one variable. Except for state license number, the combined MEDU rates for Alaska are less than the national averages.

Carrier 1030: Aetna Life and Casualty (Arizona)

Most of the rates indicating poor variable integrity are low (Table A-38). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the

other is state license number. Missing value rates are high for all seven. Erroneous value rates are one percent and higher for both versions of the medical school code.

Combined MEDU rates decline over time. Each provider type has a combined MEDU rate that exceed 25 percent for at least one variable. The combined MEDU rates for Arizona are substantially higher than the national averages.

Carrier 1040: Aetna Life and Casualty (Georgia)

Most of the rates indicating poor variable integrity are low (Table A-39). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven.

Combined MEDU rates decline over time. Combined MEDU rates are especially high for MDs. Each of the other provider types also has least one combined MEDU rate that exceeds 12 percent. The combined MEDU rates for Georgia are substantially higher than the national averages.

Carrier 1120: Aetna Life and Casualty (Hawaii)

Most of the rates indicating poor variable integrity are low (Table A-40). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Both versions of graduation year have high dubious value rates.

Combined MEDU rates decline over time. Combined MEDU rates are especially high for MDs. Each of the other provider types also has least one combined MEDU rate that exceeds one percent. Although usually lower, there are some combined MEDU rates for Hawaii that are higher than the national averages.

Carrier 1290: Aetna Life and Casualty (Nevada)

Most of the rates indicating poor variable integrity are low (Table A-41). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Both versions of the medical school code have error rates between 0.5 and 1.0 percent. The header version of graduation year has dubious value rate of 0.67 percent.

Combined MEDU rates decline over time. Each provider type has high missing values rates for state license number and at least one other combined MEDU rate exceeding one percent. Except for state license number, combined MEDU rates for Nevada are lower than the national averages.

Carrier 1360: Aetna Life and Casualty (New Mexico)

Most of the rates indicating poor variable integrity are low (Table A-42). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven.

Combined MEDU rates decline over time. Each provider type has high missing values rates for state license number and at least one other combined MEDU rate exceeding 15 percent. The combined MEDU rates for New Mexico are substantially higher than the national averages.

Carrier 1370: Aetna Life and Casualty (Oklahoma)

Most of the rates indicating poor variable integrity are low (Table A-43). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are

(both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Only the header version of the medical school code has an error rate above the standard. The dubious value rate for header version of graduation year is 0.52 percent.

Combined MEDU rates decline over time. Each provider type has high missing values rates for state license number and at least one other combined MEDU rate exceeding 8 percent. The combined MEDU rates for Oklahoma are substantially higher than the national averages.

Carrier 1380: Aetna Life and Casualty (Oregon)

Most of the rates indicating poor variable integrity are low (Table A-44). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The dubious value rate for header version of graduation year is 0.75 percent.

Combined MEDU rates decline over time. Each provider type has high missing values rates for state license number. Except for state license number, MDs tend have lower combined MEDU rates than the other provider types (but still exceeding one percent for each of the other six variables. Except for state license number, the combined MEDU rates for Oregon are substantially lower than the national averages.

Carrier 2050: Transamerica Occidental Life Insurance Co. (southern California)

Most of the rates indicating poor variable integrity are low (Table A-45). Six variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year. Missing value rates are high for all six.

Combined MEDU rates decline over time. Except for graduation year (both), combined MEDU rates tend to be higher for osteopaths than for other provider types. The combined MEDU rates for TOLIC are substantially lower than the national averages, except for graduation year.

Carrier 5130: Connecticut General Life (Idaho)

Most of the rates indicating poor variable integrity are low (Table A-46). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is group practice indicator. Missing value rates are high for all seven except group practice indicator. Erroneous value rates are high for group practice indicator. The dubious value rates for the header version of graduation year is 0.53 percent.

Combined MEDU rates decline over time. Each provider type has one combined MEDU rate that exceeds 13 percent for at least one variable. The error rate for primary specialty for dentists is relatively high at 2.47 percent. The combined MEDU rates for Idaho tend to be slightly higher than the national averages.

Carrier 5440: Connecticut General Life (Tennessee)

Most of the rates indicating poor variable integrity are low (Table A-47). Six variables, however, have combined MEDU rates 0.5 percent and above. Four of the variables are (both header and practice settings) of medical school code and graduation year; the others are birth year (setting) and state license number. Missing value rates are high for all six. The error rate for the header version of the medical school code is 0.56 percent. The dubious value rates for both versions of graduation year exceed 0.5 percent.

Combined MEDU rates decline over time. Each provider type has one combined MEDU rate that exceeds 0.5 percent for at least one variable. The combined MEDU rates for Tennessee tend to be substantially lower than the national averages.

Carrier 5535: Connecticut General Life (North Carolina)

Most of the rates indicating poor variable integrity are low (Table A-48). Six variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year. Missing value rates are high for all six. The error rate for the header version of the medical school code is 0.61 percent. The dubious value rate for the header version of graduation year is 0.64 percent.

Combined MEDU rates decline over time. Each provider type has one combined MEDU rate that exceeds 13 percent for at least one variable. Except for graduation year (both), the combined MEDU rates for North Carolina are substantially lower than the national averages.

Carrier 10072: Travelers Insurance Co. - Railroad Retirement Board - Georgia

Most of the rates indicating poor variable integrity are low (Table A-49). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Erroneous value rates are high for the header version of the medical school code and both versions of graduation year. The dubious value rate is equal to 0.71 percent for the header version of graduation year.

Ignoring 1989, for which only two practice settings were submitted to the Registry, combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 20 percent for at least one variable. Ignoring the header versions of the variables, the

combined MEDU rates for the Georgia RRB operations tend to be substantially higher than the national averages.

Carrier 10074: Travelers Insurance Co. - Railroad Retirement Board - Utah

Most of the rates indicating poor variable integrity are low (Table A-50). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. Erroneous value rates are 0.5 percent or higher for the two versions of the medical school code. The dubious value rate is equal to 0.55 percent for the header version of graduation year.

Combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 20 percent for at least one variable. Ignoring the header versions of the variables, the combined MEDU rates for the Utah RRB operations are substantially higher than the national averages.

Carrier 10230: Travelers Insurance Co. (Connecticut)

Most of the rates indicating poor variable integrity are low (Table A-51). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The dubious value rate is equal to 0.51 percent for the header version of graduation year.

Combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 20 percent for at least one variable. The combined MEDU rates for Connecticut are substantially higher than the national averages.

Carrier 10240: Travelers Insurance Co. (Minnesota)

Most of the rates indicating poor variable integrity are low (Table A-52). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven.

Combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 15 percent for at least one variable. With the exception of state license number, the combined MEDU rates for Minnesota are substantially lower than the national averages.

Carrier 10250: Travelers Insurance Co. (Mississippi)

Most of the rates indicating poor variable integrity are low (Table A-53). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The error rate is equal to 0.69 percent for the header version of the medical school code. The dubious value rate for the header version of graduation year is 0.59 percent.

Combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 15 percent for at least one variable. With the exception of state license number, the combined MEDU rates for Mississippi tend to be lower than the national averages.

Carrier 10490: Travelers Insurance Co. (Virginia)

Most of the rates indicating poor variable integrity are low (Table A-54). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The dubious value rates for both versions of graduation year are high.

Combined MEDU rates decline over time. For each provider type, combined MEDU rates exceed 17 percent for at least one variable. With the exception of state license number, the combined MEDU rates for Virginia are substantially lower than the national averages.

Carrier 11260: Missouri General American Life (Missouri)

Most of the rates indicating poor variable integrity are low (Table A-55). Seven variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The error rate is equal to 0.64 percent for the header version of the medical school code. Dubious value rates are high for both versions of graduation year.

Combined MEDU rates decline over time. Except for the excessive missing value rates for state license number, most of the combined MEDU rates for each of the provider types are between one and three percent. Except for state license number, the combined MEDU rates for Missouri are substantially lower than the national averages.

Carrier 14330: Group Health Insurance, Inc. (New York)

Most of the rates indicating poor variable integrity are low (Table A-56). Six variables, however, have combined MEDU rates 0.5 percent and above. The variables are (both header and practice settings) of birth year, medical school code, and graduation year. Missing value rates are high for all six. Both erroneous value rates and dubious value rates exceed two percent for both versions of graduation year.

Combined MEDU rates decline over time. MDs have the highest combined MEDU rates among the provider types. Each of the other provider types has a combined MEDU rate that exceeds 00.5 percent for at least one variable. The combined MEDU rates for Group Health are either significantly higher or significantly lower than the national averages.

Carrier 16360: Nationwide Insurance Co. (Ohio)

Most of the rates indicating poor variable integrity are low (Table A-57). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The error rate is equal to 0.55 percent for the header version of the medical school code.

Combined MEDU rates decline over time. Each provider type has at least one combined MEDU rate that exceeds 13 percent for at least one variable. The combined MEDU rates for Ohio are either significantly higher or significantly lower than the national averages.

Carrier 16510: Nationwide Insurance Co. (West Virginia)

Most of the rates indicating poor variable integrity are low (Table A-58). Seven variables, however, have combined MEDU rates 0.5 percent and above. Six of the variables are (both header and practice settings) of birth year, medical school code, and graduation year; the other is state license number. Missing value rates are high for all seven. The error rate exceeds five percent for both versions of the medical school code. The dubious value rates exceed 0.5 percent for both versions of graduation year.

Combined MEDU rates decline over time. Each provider type has at least one combined MEDU rate that exceeds 00.5 percent for at least one variable. Except for state license number, the combined MEDU rates for West Virginia are significantly lower than the national averages.

Carrier 21200: C & S Administrative Services (Maine)

Most of the rates indicating poor variable integrity are low (Table A-59). Ten variables, however, have combined MEDU rates 0.5 percent and above. Eight of the variables are (both

header and practice settings) of birth year, death year, medical school code, and graduation year; the other two are primary specialty code and state license number. Missing value rates are high for all ten except the two death years and primary specialty code. Erroneous value rates are high for primary specialty code. Dubious value rates are high for both versions of death year and both versions of graduation year.

Combined MEDU rates decline over time. Missing value rates exceed 40 percent for each of the provider types. MDs and dentists also have very high missing value rates for the practice setting versions of birth year, medical school code, and graduation year. The combined MEDU rates for Maine are significantly higher than the national averages.

5.0 LOGICAL INCONSISTENCIES BETWEEN DATA ELEMENTS

One of the potential problems that can occur in the UPIN database is logical inconsistency between data elements (variables). An example of a logical inconsistency is when the provider status variable indicates that the provider is a physician or an osteopath while the provider credential variable indicates that the provider is an optometrist. In this particular example, both of the variables have feasible values. It is not immediately clear, however, which of the variables is improperly coded. In some instances, the value of a third variable might help resolve the conflict, while in other cases it would be necessary for the carrier to check its hard copy files and/or directly contact the provider. (Note that logical inconsistency problems need not be limited to instances in which both variables have feasible values.)

In this chapter, logical inconsistencies between pairs of variables in the Physician Registry database are described. For each pair of variables, consistent (valid) combinations of variable values are indicated by shading in the text tables; any other combination is considered inconsistent (and unshaded). In most cases of inconsistency between two variables in which both have feasible values,[1] it is not possible to ascertain which of the variables is improperly coded when using the data at hand. Thus, in most cases, inconsistencies were noted without a determination of which variable was improperly coded.

For each pair of variables, the carriers whose records contain inconsistencies are shown in Appendix B (carriers are listed in descending order of the number of inconsistencies). The tables in Appendix B also indicate the year in which the practice setting was originally added to the Registry. This does not necessarily mean that indicated inconsistencies actually occurred in the year in which the practice setting was added. However, given that the Registry is a

[1]See Exhibit 3-3 for feasible values for each data element.

relatively new database, it is likely that most of the indicated inconsistencies occurred during the year in which the practice setting was added to the Registry.

Since Chapter 4 highlighted problems by DRIP status, deactivated records are not included in this chapter's comparisons. For comparisons between header data elements, the first practice setting, regardless of DRIP status, was initially selected for each provider (UPIN). This subset of records for the analysis of header data elements was subsequently reduced by dropping records with a de-activated DRIP status. In instances in which both setting and header versions of a variable exist, comparisons between the two versions are described in Chapter 7.

5.1 Sanction Variables

There are three types of sanction variables: (1) sanction code, (2) date when sanction took effect, and (3) length of sanction. The coding of each variable was compared to values of the other two variables, yielding three pair-wise comparisons. Additionally, there are setting and header versions of all three variables. As noted above, the comparisons between the three variables, in this chapter, did *not* include setting versus header versions.

The distribution of the settings version of sanction code by date of sanction is shown in Table 5-1. The shaded cells indicate consistent (valid) combinations of values of the two variables. For this pair of variables, there are three valid combinations. The first is that when there is no sanction code, a date of sanction should not be indicated. In all but eight cases out of 1,679,079 in which no sanction code was indicated, no sanction date was indicated.

For the analysis of feasible sanction codes, the list was divided into codes *A-R* and *U*. Code *U* is a legitimate code in HCFA's list of sanction codes and indicates that the provider is sanctioned but that the type or reason for the sanction is not known. In only about two-thirds of the cases in which codes *A-R* were used, a feasible date was indicated. In about one-third of the cases in which codes *A-R* were used, no sanction date was indicated – an inconsistency. That is, there were 621 cases in which the sanction code indicated the provider was sanctioned while the sanction date could be interpreted to mean the provider was *not* sanctioned.

TABLE 5-1

DISTRIBUTION OF SANCTION CODES BY CODING ON DATE OF SANCTION (SETTING)

	CODING ON DATE OF SANCTION				
		INFEASIBLE			
Sanction Code	No Date	Year	Month	Feasible Date	Total
No Code	1,679,071 (100%)	1 (0.0%)	1 (0.0%)	6 (0.0%)	1,679,079
Feasible Codes: A-R	621 (32.7%)	68 (3.6%)	36 (1.9%)	1,173 (61.8%)	1,898
Feasible Code: U (Unknown)	1,022 (90.2%)	29 (2.6%)	0 (0.0%)	82 (7.2%)	1,133
Total	1,680,714	98	37	1,261	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

No sanction date was present for about 90 percent of the 1,133 cases in which code *U* was used. This is another instance in which one variable indicated that the provider was sanctioned while the other variable failed to corroborate it. Considering all 3,031 cases in which a sanction code was used, only 41.4 percent of the cases had a feasible sanction date.

Table B-1 in Appendix B indicates that two carriers accounted for nearly 90 percent of the records containing inconsistencies: North Dakota BC/BS operations for North and South Dakota and Group Health Insurance of New York. In the case of North Dakota BC/BS, about 20 percent of its active practice settings have inconsistent sanction codes and dates while for Group Health Insurance, it is about 7 percent. The vast majority of the inconsistent cases for not only these two carriers, but most carriers, apparently occurred during the implementation phase (1989) of the Registry (Table B-1). The major exception is RRB operations in Georgia, which had 55 cases in 1990-1 and another 16 in 1992-3.

The header versions of sanction code and sanction date were also compared to each other (Table 5-2). The distribution of consistent and inconsistent combinations in Table 5-2 is about the same as in Table 5-1. Of the 1,114 providers that still have active practice settings and that have a feasible sanction code, only 654 (58.7 percent) have feasible dates. Because changes to the header versions of sanction variable can originate with any carrier having a setting record for the provider, we did not associate carriers with the inconsistent occurrences – that is, appendix tables for comparisons of the header versions of the sanction variables were *not* created.

The distribution of the setting versions of sanction codes by length of sanction is shown in Table 5-3. For length of sanction, there are two types of feasible values, one indicating the actual number of years sanctioned and the other showing a value of 99 which indicates an indefinite sanction. For all but 35 instances in which there was no sanction code, no length of sanction was indicated. For sanction codes *A-R*, about 61 percent of the cases had feasible lengths of sanction while about a third of the cases did not have a length of sanction value

TABLE 5-2

DISTRIBUTION OF SANCTION CODES BY CODING ON DATE OF SANCTION (HEADER)

	CODING ON DATE OF SANCTION				
		INFEASIBLE			
Sanction Code	No Date	Year	Month	Feasible Date	Total
No Code	584,463 (100.0%)	18 (0.0%)	9 (0.0%)	143 (0.0%)	584,633
Invalid Code	1 (100.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1
Feasible Code: A-R	237 (26.3%)	38 (4.2%)	9 (1.0%)	619 (68.6%)	903
Feasible Code: U (Unknown)	170 (80.6%)	6 (2.8%)	0.0 (0.0%)	35 (16.6%)	211
Total	584,871	62	18	797	585,748

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.
The first practice setting, regardless of DRIP status, was initially selected for each provider record (UPIN).
This subset was subsequently reduced by dropping records with a de-activated DRIP status.

SOURCE: Physician Registry, July 1993.

TABLE 5-3

DISTRIBUTION OF SANCTION CODES BY CODING ON LENGTH OF SANCTION (SETTING)

	CODING ON LENGTH OF SANCTION				
Sanction Code	No Code	Feasible	Dubious	Feasible (Indefinite)	Total
No Code	1,679,044 (100%)	28 (0.0%)	5 (0.0%)	2 (0.0%)	1,679,079
Feasible Code: A-R	663 (34.9%)	781 (41.2%)	73 (3.9%)	381 (20.1%)	1,898
Feasible Code: U (Unknown)	1,049 (92.6%)	28 (2.5%)	4 (0.4%)	52 (4.6%)	1,133
TOTAL	1,680,756	837	82	435	1,682,110

Note: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

indicated. For sanction code *U*, over 90 percent did not have a length of sanction value. Between the 3,031 instances of a feasible sanction code, only 1,242 cases (41 percent) had feasible lengths of sanction. Table 5-3 supports the inferences suggested by Table 5-1, that is, the consistency of sanction variable coding is very poor, especially in instances in which a feasible sanction code is indicated.

Table B-2 indicates that North Dakota BC/BS and Group Health Insurance of New York again had the majority of cases with inconsistent coding, with most cases apparently occurring during the implementation year (1989). The distribution of the header versions of sanction codes by length of sanction is shown in Table 5-4 and is very similar to Table 5-3's distribution of codes.

Table 5-5 shows the frequencies for the setting version of the sanction length by date of sanction. Unlike the previous tables, a feasible length of sanction was usually accompanied by a feasible date of sanction as indicated by the 90+ percent row percentage values in the shaded cells. The count of 214 inconsistent combinations is considerably lower than the total number of inconsistent combinations indicated in Tables 5-1 or 5-3. The six carriers (Table B-3) with the highest number of inconsistencies were RRB Georgia operations, South Carolina BC/BS, Blue Shield of Western New York, Illinois BC/BS, Kansas BS, and Pennsylvania BS (New Jersey). Note, however, the *total* number of inconsistencies is low by comparison to counts indicated in Tables B-1 and B-2, and the percentage of practice settings affected is usually less than 0.1 percent. Table 5-6 shows the header version of the distribution of length of sanction values by date of sanction and is similar to the distribution shown in Table 5-5.[2]

Taken together, Tables 5-1, 5-3, and 5-5 suggest that during the implementation phase of the Registry, if a carrier knew that a provider was sanctioned it often did not know the date of sanction and the length of sanction. This would account both for the high number of

[2]It is possible that taken together sanction date and sanction length indicate that a provider is no longer sanctioned. Technically, however, these providers are still sanctioned until they apply for reinstatement, and then the sanction information should be updated to indicate that the provider is not sanctioned.

TABLE 5-4

DISTRIBUTION OF SANCTION CODES BY CODING ON LENGTH OF SANCTION (HEADER)

	CODING ON LENGTH OF SANCTION				
Sanction Code	No Code	Feasible	Dubious	Feasible (Indefinite)	Total
No Code	584,451 (100%)	115 (0.0%)	13 (0.0%)	54 (0.0%)	584,633
Invalid Code	1 (100%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1
Feasible Code: A-R	260 (28.8%)	394 (43.6%)	43 (4.8%)	206 (22.8%)	903
Feasible Code: U (Unknown)	176 (83.4%)	13 (6.2%)	2 (1.0%)	20 (9.5%)	211
Total	584,888	522	58	280	585,748

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.
The first practice setting, regardless of DRIP status, was initially selected for each provider record (UPIN).
This subset was subsequently reduced by dropping records with a de-activated DRIP status.

SOURCE: Physician Registry, July 1993.

TABLE 5-5

DISTRIBUTION OF CODING ON LENGTH OF SANCTION BY CODING ON DATE OF SANCTION (SETTING)

	CODING ON DATE OF SANCTION				
Length of Sanction	No Date	INFEASIBLE Year	INFEASIBLE Month	Feasible Date	Total
No Code	1,680,680 (100%)	26 (0.0%)	10 (0.0%)	40 (0.0%)	1,680,756
Feasible	26 (3.1%)	0 (0.0%)	21 (2.5%)	790 (94.4%)	837
Dubious	5 (6.1%)	72 (87.8%)	0 (0.0%)	5 (6.1%)	82
Feasible (Indefinite)	3 (0.7%)	0 (0.0%)	6 (1.4%)	426 (97.9%)	435
Total	1,680,714	98	37	1,261	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

TABLE 5-6

DISTRIBUTION OF CODING ON LENGTH OF SANCTION BY CODING ON DATE OF SANCTION (HEADER)

	CODING ON DATE OF SANCTION				
Length of Sanction	No Date	INFEASIBLE Year	INFEASIBLE Month	Feasible Date	Total
No Code	584,837 (100%)	11 (0.0%)	4 (0.0%)	36 (0.0%)	584,888
Feasible	22 (4.2%)	1 (0.2%)	10 (1.9%)	489 (93.7%)	522
Dubious	3 (5.2%)	50 (86.2%)	0 (0.0%)	5 (8.6%)	58
Feasible (Indefinite)	9 (3.2%)	0 (0.0%)	4 (1.4%)	267 (95.4%)	280
Total	584,871	62	18	797	585,748

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.
The first practice setting, regardless of DRIP status, was initially selected for each provider record (UPIN).
This subset was subsequently reduced by dropping records with a de-activated DRIP status.

SOURCE: Physician Registry, July 1993.

inconsistencies occurring in practice settings added to the Registry during 1989 and for the lower number of inconsistencies occurring in Table 5-5.

5.2 Provider Credentials, Provider Status, and Specialty

In this section, we examine consistency in the coding of provider credentials, provider status, and specialty. Because header versions of these variables do not exist, only setting level observations were examined. The assignment of HCFA specialty codes to specialty types is listed in Exhibit 5-1.

Table 5-7 shows provider credentials by provider status. There are seven feasible provider credential codes and two feasible provider status codes. In almost all cases, for a given provider credential code, the provider status was consistent with it, as evidenced by the near 100 percent values in the shaded cells. (Note that, due to rounding, none of the cells showing 100% actually does contain all cases of a given credential.) On a percentage basis, optometrists, podiatrists, and doctors of dental surgery most often had an inconsistent recorded provider status. Over 95 percent of cases were due to the two RRB carriers and were of recent origin (Table B-4).

Combinations of provider credentials and *primary* specialty codes are shown in Table 5-8. Most providers (99.8 percent) that have a medical doctor credential also had a specified *medical doctor* specialty. Of those that did not, most had an *unknown* or *miscellaneous* specialty code. The *unknown* and *miscellaneous* specialty codes are technically feasible but are most likely indicative of cases where the physician's specialty was not classifiable into one of the available codes. Also, the *miscellaneous* code (49) is no longer a valid physician code; however, since it was a valid physician code when the Registry was established, it is indicated as valid in the table. Of the other specified specialties, osteopathy was the most often recorded (1,081 setting records) when a provider had a medical doctor credential indicated.

EXHIBIT 5-1

PROVIDER SPECIALTY TYPES IMPLIED BY THE HCFA SPECIALTY CODE

SPECIALTY	SPECIALTY CODE
Medical doctor	01, 02, 03, 04, 05, 06, 07, 08, 10, 11, 13, 14, 16, 18, 20, 22, 24, 25, 26, 28, 29, 30, 33, 34, 36, 37, 38, 39, 40, 44, 46, 66, 70, 76, 77, 78, 79, 81, 82, 83, 84, 86, 90, 91, 92, 93, 94
Osteopathy	09, 12, 15, 17 ,21, 23, 27, 31, 32
Oral surgery	19
Maxillofacial surgery	85
Chiropractic	35
Podiatry	48
Optometry	41
Miscellaneous	49
Unknown	99

TABLE 5-7

PROVIDER CREDENTIALS BY PROVIDER STATUS (SETTING)

	PROVIDER STATUS			
Provider Credentials	Invalid	Medical Doctor/ Osteopath	All Other Providers	Total
Medical Doctor	1 (0.0%)	1,424,210 (100%)	13 (0.0%)	1,424,224
Doctor of Osteopathy	0 (0.0%)	53,963 (100%)	1 (0.0%)	53,964
Doctor of Dental Medicine	0 (0.0%)	1 (0.0%)	10,143 (100%)	10,144
Doctor of Dental Surgery	0 (0.0%)	13 (0.0%)	39,392 (100%)	39,405
Chiropractor	0 (0.0%)	3 (0.0%)	67,684 (100%)	67,687
Podiatrist	0 (0.0%)	21 (0.1%)	36,702 (99.9%)	36,723
Doctor of Optometry	0 (0.0%)	56 (0.1%)	49,907 (99.9%)	49,963
Total	1	1,478,267	203,842	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

TABLE 5-8

PROVIDER CREDENTIALS BY PRIMARY SPECIALTY (SETTING)

	PRIMARY SPECIALTY									
Provider Credentials	Medical Doctor	Osteopathy	Oral Surgery	Maxillofacial Surgery	Chiropractic	Podiatry	Optometry	Unknown	Misc.	Total
Medical Doctor	1,421,178 (99.8%)	1,081 (0.1%)	89 (0.0%)	140 (0.0%)	137 (0.0%)	14 (0.0%)	58 (0.0%)	250 (0.0%)	1,277 (0.1%)	1,424,224
Doctor of Osteopathy	50,133 (92.9%)	3,466 (6.4%)	1 (0.0%)	1 (0.0%)	4 (0.0%)	1 (0.0%)	9 (0.0%)	17 (0.0%)	332 (0.6%)	53,964
Doctor of Dental Medicine	58 (0.6%)	0 (0.0%)	9,697 (95.6%)	104 (1.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	99 (1.0%)	185 (1.8%)	10,144
Doctor of Dental Surgery	251 (0.6%)	0 (0.0%)	37,993 (96.4%)	888 (2.3%)	0 (0.0%)	1 (0.0%)	6 (0.0%)	161 (0.4%)	105 (0.3%)	39,405
Chiropractor	64 (0.1%)	3 (0.0%)	2 (0.0%)	0 (0.0%)	67,602 (99.9%)	7 (0.0%)	6 (0.0%)	3 (0.0%)	0 (0.0%)	67,687
Podiatrist	348 (1.0%)	4 (0.0%)	5 (0.0%)	0 (0.0%)	4 (0.0%)	36,358 (99.0%)	1 (0.0%)	0 (0.0%)	3 (0.0%)	36,723
Doctor of Optometry	485 (1.0%)	3 (0.0%)	4 (0.0%)	0 (0.0%)	0 (0.0%)	5 (0.0%)	49,464 (99.0%)	2 (0.0%)	0 (0.0%)	49,963
Total	1,472,517	4,557	47,791	1,133	67,747	36,387	49,544	532	1902	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

Because HCFA directed carriers to phase out osteopathic codes during the early 1990s, osteopaths should have a specialty code indicative of a medical specialty. However, for the purposes of this table, osteopaths were allowed to have osteopathic specialty codes; only 6.4 percent of practice settings that have the osteopath credential still have an osteopathic specialty code. Aside from medical doctors, osteopaths are the only providers that we allowed to have *unknown* or *miscellaneous* specialty codes.

Most dentists (over 97 percent) had oral surgery or maxillofacial surgery as their listed primary specialty; however, medical doctor specialties occurred in a relatively small but significant number of cases. Similarly, chiropractors, podiatrists, and optometrists almost always had a primary specialty code consistent with their provider credentials. In inconsistent cases, a medical doctor specialty was usually indicated for them.

The carriers with the most inconsistencies between provider credentials and primary specialty code are indicated in Table B-5. California BS had the highest number of active practice settings with 1,142 inconsistent values, followed by Pennsylvania BS (Pennsylvania) at 414, C&S Administrative Services (Massachusetts) at 345, and Blue Shield of Western New York with 174. Along with Rhode Island BS, these four carriers also had the highest percentage of practices settings with inconsistencies. The inconsistencies do not always decline with the year the setting was added to the Registry.

While a provider must have a primary specialty, a provider might not have a secondary specialty. Thus, the lack of a secondary specialty is not considered to be inconsistent with any given provider credential and, hence, all of the cells in the column indicating no secondary specialty in Table 5-9 are shaded. Where a secondary specialty is given, most codes were consistent with the indicated credential. As in Table 5-8, many of the inconsistencies occurred when a medical doctor specialty was indicated in conjunction with dental, chiropractor, podiatrist, and optometrist credentials. Unlike Table 5-8, there were also a number of cases in which a medical doctor credential was indicated but the indicated specialty was that for a non-

TABLE 5-9

PROVIDER CREDENTIALS BY SECONDARY SPECIALTY (SETTING)

	SECONDARY SPECIALTY										
Provider Credentials	No Secondary Specialty Indicated	Medical Doctor	Osteopath	Oral Surgery	Maxillofacial Surgery	Chiropractic	Podiatry	Optometry	Unknown	Misc.	Total
Medical Doctor	1,338,312 (94.0%)	85,523 (6.0%)	148 (0.0%)	32 (0.0%)	55 (0.0%)	3 (0.0%)	12 (0.0%)	63 (0.0%)	39 (0.0%)	37 (0.0%)	1,424,224
Doctor of Osteopathy	50,257 (93.1%)	3,472 (6.4%)	228 (0.4%)	0 (0.0%)	2 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)	2 (0.0%)	1 (0.0%)	53,964
Doctor of Dental Medicine	9,867 (97.3%)	8 (0.1%)	0 (0.0%)	258 (2.5%)	10 (0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	10,144
Doctor of Dental Surgery	38,362 (97.4%)	72 (0.2%)	0 (0.0%)	855 (2.2%)	113 (0.3%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	2 (0.0%)	0 (0.0%)	39,405
Chiropractor	64,108 (94.7%)	56 (0.1%)	2 (0.0%)	2 (0.0%)	0 (0.0%)	3,517 (5.2%)	0 (0.0%)	0 (0.0%)	2 (0.0%)	0 (0.0%)	67,687
Podiatrist	35,850 (97.6%)	44 (0.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (0.0%)	827 (2.3%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	36,723
Doctor of Optometry	47,758 (95.6%)	116 (0.2%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2,087 (4.2%)	1 (0.0%)	0 (0.0%)	49,963
Total	1,584,514	89,291	378	1,148	180	3,522	841	2,151	47	38	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

medical specialty. Iowa BS with 107 and Aetna (New Mexico) with 80 had the highest number and highest percentage of practice settings with coding inconsistencies between credential and secondary specialty (Table B-6).

Combinations of primary specialty code and provider status are indicated in Table 5-10. Most of the primary specialty types were consistent with their provider status. For all but three of the specialty codes, more than 99 percent of cases indicated a consistent provider status value. About 12.4 percent of practice settings with a maxillofacial surgery specialty had a medical doctor provider status indicated. Only half of the *unknown* specialty codes were correctly associated with the medical doctor or osteopath provider status. Nearly 85 percent of the *miscellaneous* specialty codes were correctly associated with medical doctor provider status. Four carriers had more than 100 inconsistencies between primary specialty code and provider status (Table B-7): California BS, Pennsylvania BS (Pennsylvania) Blue Shield of Western New York, and the RRB operations in Georgia. The four carriers with highest percent of inconsistencies were Blue Shield of Western New York, Rhode Island BS, California BS, and Pennsylvania BS (Pennsylvania).

Combinations of secondary specialty code and provider status are shown in Table 5-11. For all but two secondary specialty codes, more than 95 percent of specialty codes had a consistent provider status. The two falling short of 90 percent were maxillofacial surgery and *unknown* specialty. Because of the high percentage of practice settings that did not indicate a secondary specialty code, the total number of inconsistencies is much lower than when *primary* specialty code was compared to provider status. Iowa BS and Aetna (New Mexico) had the highest number of inconsistencies and the highest percentage of practice settings with inconsistencies (Table B-8).

TABLE 5-10

PRIMARY SPECIALTY BY PROVIDER STATUS (SETTING)

		PROVIDER STATUS		
Primary Specialty	Invalid	Medical Doctor/ Osteopath	All Other Providers	Total
Medical Doctor	1 (0.0%)	1,471,299 (99.9%)	1,217 (0.1%)	1,472,517
Osteopathy	0 (0.0%)	4,547 (99.8%)	10 (0.2%)	4,557
Oral Surgery	0 (0.0%)	102 (0.2%)	47,689 (99.8%)	47,791
Maxillofacial Surgery	0 (0.0%)	141 (12.4%)	992 (87.6%)	1,133
Chiropractic	0 (0.0%)	144 (0.2%)	67,603 (99.8%)	67,747
Podiatry	0 (0.0%)	35 (0.1%)	36,352 (99.9%)	36,387
Optometry	0 (0.0%)	122 (0.3%)	49,422 (99.8%)	49,544
Unknown	0 (0.0%)	267 (50.2%)	265 (49.8%)	532
Misc.	0 (0.0%)	1,610 (84.7%)	292 (15.4%)	1,902
Total	1	1,478,267	203,842	1,682,110

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

TABLE 5-11

SECONDARY SPECIALTY BY PROVIDER STATUS (SETTING)

		PROVIDER STATUS		
Secondary Specialty	Invalid	Medical Doctor/ Osteopath	All Other Providers	Total
No Secondary Specialty Indicated	1 (0.0%)	1,388,641 (87.6%)	195,872 (12.4%)	1,584,514
Medical Doctor	0 (0.0%)	88,996 (99.7%)	295 (0.3%)	89,291
Doctor of Osteopathy	0 (0.0%)	376 (99.5%)	2 (0.5%)	378
Oral Surgery	0 (0.0%)	33 (2.9%)	1,115 (97.1%)	1,148
Maxillofacial Surgery	0 (0.0%)	58 (32.2%)	122 (67.8%)	180
Chiropractic	0 (0.0%)	3 (0.1%)	3,519 (99.9%)	3,522
Podiatry	0 (0.0%)	14 (1.7%)	827 (98.3%)	841
Optometry	0 (0.0%)	67 (3.1%)	2,084 (96.9%)	2,151
Unknown	0 (0.0%)	41 (87.2%)	6 (12.8%)	47
Misc.	0 (0.0%)	38 (100%)	0 (0.0%)	38
Total	1	1,478,267	203,842	1,682,110

Notes: Row percentages in parentheses.

Percentages may add to more than 100% due to rounding.

Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

5.3 School Type, Provider Credentials, Provider Status, Specialty, and Graduation Year

This section compares the type of school from which the provider graduated to provider credentials, provider status, specialty code, and graduation year. (We use the generic term *school code* instead of *medical school code* (as used in the Registry's instructions to carriers) to avoid confusion when referring to actual medical schools.) The text tables *exclude* observations with missing or invalid school codes. The tables in Appendix B, however, *include* such observations, which affect both the counts in the first four columns and the percentages in the last two columns. The difference between the two methods of counting the schools results in 304,534 fewer observations (mostly due to a missing school code) in the text tables than in Appendix B tables. Note that according to the *UPIN Carrier Record Layout*, when the school was not known, the school field is supposed to be filled with blanks. However, the instructions to carriers were not explicit on how carriers were to code the school field when they knew the identity of the school but there was not an assigned code available for it (*Medicare Carriers Manual*, Part 4, Transmittal #2, October 1989).

To determine which types of providers could legitimately indicate education at a foreign school, professional associations were contacted. Association representatives indicated that graduates of the following types of foreign schools are eligible to be licensed in the United States: medical doctors, dentists, chiropractors, and optometrists.

School type is compared to provider credentials in Table 5-12. Most practice settings that had a valid *medical* school code also had medical doctor credentials indicated. Most of the inconsistencies when a medical school code was indicated were when an osteopath credential was indicated (3,231 cases). Of the practice settings that had a valid osteopathic school code, provider credentials were not consistent in 27.1 percent of the cases; invariably a medical doctor credential was indicated instead. For those practice settings indicating dental school, 99.9 percent had a DDM or DDS credential indicated. Except when a podiatric school was

TABLE 5-12

SCHOOL TYPES BY PROVIDER CREDENTIALS (SETTING)

	PROVIDER CREDENTIALS							
			DENTIST					
School Types	Medical Doctor	Osteopath	DDM	DDS	Chiropractor	Podiatrist	Optometrist	Total
Medical	870,171 (99.6%)	3,231 (0.4%)	1 (0.0%)	2 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)	873,407
Osteopathic	14,913 (27.1%)	40,138 (72.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	55,051
Dental	0 (0.0%)	0 (0.0%)	8,632 (18.4%)	38,287 (81.5%)	11 (0.0%)	35 (0.1%)	27 (0.1%)	46,992
Chiropractic	0 (0.0%)	0 (0.0%)	2 (0.0%)	27 (0.0%)	62,495 (99.8%)	52 (0.1%)	27 (0.0%)	62,603
Podiatric	461 (1.3%)	1 (0.0%)	9 (0.0%)	8 (0.0%)	43 (0.1%)	34,831 (98.4%)	34 (0.1%)	35,387
Optometric	1 (0.0%)	0 (0.0%)	2 (0.0%)	12 (0.0%)	31 (0.1%)	94 (0.2%)	45,518 (99.7%)	45,658
Foreign	256,600 (99.3%)	1,878 (0.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	258,478
Total	1,142,146	45,248	8,646	38,336	62,580	35,013	45,607	1,377,576

Notes: All foreign medical schools are medical or osteopathic.
Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

indicated, the other school types had over 99 percent of their cases consistent with the indicated credential. Table 5-12 indicates that the only providers aside from medical doctors that had a foreign school indicated were osteopaths. According to the American Osteopathic Association, however, foreign-trained osteopaths are not eligible to practice in the United States.

When considering missing and invalid school codes, four carriers had over 20,000 inconsistencies between school type and credential with most of the inconsistencies due to missing or erroneous school codes (Table B-9). The four were the two RRB carriers, Empire BC/BS, and C&S Administrative Services (Massachusetts). As a percent of carrier's active practice settings, the four carriers with the highest percentage of inconsistencies were South Carolina BC/BS, Aetna (Georgia), Arkansas BC/BS, and C&S Administrative Services (Massachusetts).

When school type is compared to provider status (Table 5-13), the total number of inconsistencies is lower than in Table 5-12, mostly because osteopaths are combined with medical doctors in the coding for provider status. The single largest problem was when a podiatric school was indicated: there were 481 cases in which the medical doctor/osteopath provider status was indicated instead of *other* provider. The carriers with the highest number of inconsistencies were the two RRB carriers, Empire BC/BS, and C&S Administrative Services (Massachusetts) – all had over 20,000 cases (Table B-10). As a percent of carrier's active practice settings, the four carriers with the highest percentage of inconsistencies, more than 35 percent, are South Carolina BC/BS, Aetna (Georgia), Arkansas BC/BS (Arkansas), and C&S Administrative Services (Massachusetts).

School type and primary specialty combinations are shown in Table 5-14. For any particular school type indicated, the inconsistency rate was no higher than 2.6 percent (medical doctor specialty accounted for most of the inconsistent specialty codes in conjunction with podiatric school code). Other problems included osteopath specialty indicated in conjunction with medical school (but only 0.1 percent of cases with medical school code indicated), medical

TABLE 5-13

SCHOOL TYPE BY PROVIDER STATUS

		PROVIDER STATUS		
School Type	Invalid	Medical Doctor Osteopath	All Other Providers	Total
Medical	1 (0.0%)	873,396 (100%)	10 (0.0%)	873,407
Osteopathic	0 (0.0%)	55,049 (100%)	2 (0.0%)	55,051
Dental	0 (0.0%)	13 (0.0%)	46,979 (100%)	46,992
Chiropractic	0 (0.0%)	3 (0.0%)	62,600 (100%)	62,603
Podiatric	0 (0.0%)	481 (1.4%)	34,906 (98.6%)	35,387
Optometric	0 (0.0%)	54 (0.1%)	45,604 (99.9%)	45,658
Foreign	0 (0.0%)	258,475 (100%)	3 (0.0%)	258,478
Total	1	1,187,471	190,104	1,377,576

Notes: All foreign medical schools are medical or osteopathic.
Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

TABLE 5-14

SCHOOL TYPES BY PRIMARY SPECIALTY (SETTING)

School Types	PRIMARY SPECIALTY									
	Medical Doctor	Osteopathy	Oral Surgery	Maxillofacial Surgery	Chiropractic	Podiatry	Optometry	Unknown	Misc.	Total
Medical	870,686 (99.7%)	1,187 (0.1%)	67 (0.0%)	116 (0.0%)	97 (0.0%)	6 (0.0%)	47 (0.0%)	190 (0.0%)	1,011 (0.1%)	873,407
Osteopathic	53,217 (96.7%)	1,605 (2.9%)	0 (0.0%)	1 (0.0%)	3 (0.0%)	2 (0.0%)	4 (0.0%)	13 (0.0%)	206 (0.4%)	55,051
Dental	302 (0.6%)	0 (0.0%)	45,097 (96.0%)	975 (2.1%)	11 (0.0%)	33 (0.1%)	28 (0.1%)	260 (0.6%)	286 (0.6%)	46,992
Chiropractic	54 (0.1%)	3 (0.0%)	31 (0.1%)	0 (0.0%)	62,421 (99.7%)	58 (0.1%)	33 (0.1%)	3 (0.0%)	0 (0.0%)	62,603
Podiatric	805 (2.3%)	4 (0.0%)	15 (0.0%)	0 (0.0%)	46 (0.1%)	34,478 (97.4%)	35 (0.1%)	0 (0.0%)	4 (0.0%)	35,387
Optometric	485 (1.1%)	3 (0.0%)	13 (0.0%)	0 (0.0%)	29 (0.1%)	98 (0.2%)	45,028 (98.6%)	2 (0.0%)	0 (0.0%)	45,658
Foreign	257,871 (99.8%)	285 (0.1%)	12 (0.0%)	19 (0.0%)	31 (0.0%)	7 (0.0%)	7 (0.0%)	36 (0.0%)	210 (0.1%)	258,478
Total	1,183,420	3,087	45,235	1,111	62,638	34,682	45,182	504	1,717	1,377,576

Notes All foreign medical schools are medical or osteopathic.
Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values

SOURCE: Physician Registry, July 1993.

doctor specialty indicated in conjunction with optometric school, medical doctor specialty indicated in conjunction with dental school, and maxillofacial surgery specialty in conjunction with medical school. When considering missing and invalid school codes, the two RR carriers, Empire BC/BS, and C&S Administrative Services (Massachusetts) had the highest number of inconsistencies between school type and primary specialty code with over 20,000 each (Table B-11). The four carriers that had more than 35 percent of their active practice settings with inconsistencies were South Carolina BC/BS, Aetna (Georgia), Arkansas BC/BS (Arkansas), and C&S Administrative Services (Massachusetts).

School type and *secondary* specialty combinations are shown in Table 5-15. Although the absolute number of inconsistencies is lower than in the previous table, for any particular school type indicated, overall inconsistency percentage rates were higher than for school type and primary specialty. (The reason is that the observations for which the secondary specialty code was not indicated were not included in constructing the table.) Notable inconsistencies included osteopath specialty indicated in conjunction with medical school, medical doctor specialty indicated in conjunction with optometric school, medical doctor specialty in conjunction with dental school, medical doctor specialty in conjunction with chiropractic school, medical doctor specialty in conjunction with podiatric school and maxillofacial surgery specialty in conjunction with medical school. Aetna (different states) had the three highest numbers and percentage rates of inconsistencies between school type and secondary specialty although the problems declined after 1989 (Table B-12). Note that observations for which the secondary specialty code was not indicated were not included in either the numerator nor denominator for Table B-12.

The year of graduation from a medical school was compared to the establishment (open) date and the closure date (if closed) of the indicated medical school. The setting versions of the two variables are presented in Table 5-16 and the header versions of the two variables are presented in Table 5-17. Only medical schools were included in the tables because

TABLE 5-15

SCHOOL TYPES BY SECONDARY SPECIALTY (SETTING)

	SECONDARY SPECIALTY									
School Types	Medical Doctor	Osteopathy	Oral Surgery	Maxillofacial Surgery	Chiropractic	Podiatry	Optometry	Unknown	Misc.	Total
Medical	53,663 (99.6%)	87 (0.2%)	27 (0.1%)	42 (0.1%)	0 (0.0%)	8 (0.0%)	20 (0.0%)	33 (0.1%)	18 (0.0%)	53,898
Osteopathic	3,306 (94.1%)	200 (5.7%)	0 (0.0%)	3 (0.1%)	1 (0.0%)	1 (0.0%)	0 (0.0%)	3 (0.1%)	1 (0.0%)	3,515
Dental	54 (6.0%)	0 (0.0%)	725 (81.0%)	111 (12.4%)	0 (0.0%)	1 (0.1%)	1 (0.1%)	3 (0.3%)	0 (0.0%)	895
Chiropractic	35 (1.4%)	2 (0.1%)	3 (0.1%)	0 (0.0%)	2,396 (98.3%)	0 (0.0%)	0 (0.0%)	2 (0.1%)	0 (0.0%)	2,438
Podiatric	18 (2.7%)	0 (0.0%)	4 (0.6%)	0 (0.0%)	2 (0.3%)	653 (96.0%)	3 (0.4%)	0 (0.0%)	0 (0.0%)	680
Optometric	102 (6.6%)	0 (0.0%)	2 (0.1%)	0 (0.0%)	7 (0.5%)	7 (0.5%)	1,428 (92.3%)	1 (0.1%)	0 (0.0%)	1,547
Foreign	12,132 (99.6%)	21 (0.2%)	1 (0.0%)	6 (0.1%)	0 (0.0%)	4 (0.0%)	5 (0.0%)	4 (0.0%)	4 (0.0%)	12,177
Total	69,310	310	762	162	2,406	674	1,457	46	23	75,150

Notes: All foreign medical schools are medical or osteopathic.
Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or vlaues.

SOURCE: Physician Registry, July 1993.

upin\tables\TAB5-15.XLS\sjb

TABLE 5-16

GRADUATION YEAR COMPARED TO MEDICAL SCHOOL ESTABLISHMENT DATE AND CLOSURE DATE (SETTING)

Coding on Graduation Year	GRADUATION YEAR: Not Within Open and Close Date	GRADUATION YEAR: Within Open and Close Date	Total
Dubious	174 (3.8%)	4,376 (96.2%)	4,550
Feasible	13,316 (1.6)%	809,613 (98.4)%	822,929
Total	13,490	813,989	827,479

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

TABLE 5-17

GRADUATION YEAR COMPARED TO MEDICAL SCHOOL ESTABLISHMENT DATE AND CLOSURE DATE (HEADER)

Coding on Graduation Year	GRADUATION YEAR: Not Within Open and Close Date	GRADUATION YEAR: Within Open and Close Date	Total
Dubious	118 (3.5%)	3,257 (96.5%)	3,375
Feasible	5,888 (1.8%)	323,230 (98.2%)	329,118
Total	6,006	325,487	332,493

Notes: Row percentages in parentheses.
Percentages may add to more than 100% due to rounding.
Shading indicates valid combinations of codes, types of codes, or values.

SOURCE: Physician Registry, July 1993.

it was not possible to obtain establishment and closure dates of defunct non-medical schools. We were not able to find the establishment and closure date of one defunct school believed to have been a medical school (Washington College of Physicians and Surgeons [possibly located in Seattle, Washington]) and thus did not include providers which indicated attendance at the school in either table.

Of the practice settings with feasible graduation dates, 13,316 (1.6 percent) of the graduation dates either preceded the establishment date or came after the closure date (if closed) of the medical school (Table 5-16). For those practice settings with dubious graduation dates (1930-39), 3.8 percent were not consistent with the open or closure date of the school. The carriers with the highest number of inconsistencies were the two RRB carriers with over 2,000 apiece (Table B-13). Nationwide (WV) and the two RRB carriers had the highest percentages of inconsistencies with graduation date and open/closure dates of medical schools. (Practice settings with dubious graduation dates that are consistent with the open/closure dates are *not* included in Table B-13.)

For the header versions of graduation dates and open/closure dates of medical schools, 1.8 percent of feasible graduation dates were inconsistent with the open/closure dates of medical schools (Table 5-17). Of the dubious graduation dates, 3.5 percent were not consistent with the open or closure date of the school. A table indicating the carriers with the highest problem rates for the header versions of graduation dates and open/closure dates of medical schools was not created because the source of the information for the two variables may come from more than one carrier.

5.4 Summary

Most of the problems of inconsistent coding between pairs of variables seems to be due to missing information on one of the two. This is especially evident with regard to sanction

dates, length of sanction, and school code. In turn, most of the missing values seem to be associated with practice settings added to the Registry during 1989. With regard to school code, it was not a required data element during the implementation phase. After the implementation phase, it appears that carriers had difficulty obtaining the school code for practice settings already in the Physician Registry.

There appears to be a different type of problem with regard to sanction variables. If the conjecture is correct that carriers usually knew if a provider was sanctioned prior to the Registry's implementation, then it seems that they did not know the sanction date nor the length of sanction. (The codes currently used by the Registry that can be used for sanction date and length of sanction do not allow differentiation between *no sanction* and *sanctioned, but missing sanction date or length of sanction information.*) To reduce errors and inconsistencies, HCFA is considering centralizing the entry of sanction information.

Aside from inconsistencies involving school codes and sanction variables, the most significant problems seem to be with specialty codes being inconsistent with provider credentials, provider status, and school type. Corrections, if undertaken by a small number of carriers, would result in significant amelioration of these inconsistencies. Note that specialty codes 49[3] and 99 should be reserved for medical doctors or osteopaths with a nonclassifiable specialty. Unless it is assumed that the credential is in error, this suggests that carrier training in specialty coding may need refreshing. In conjunction with medical school codes, the inconsistency rates of graduation year (both feasible and dubious) with medical school open/close dates exceeded 1.5 percent.

Despite the above problems, in the overwhelming majority of cases (as evidenced by the 95+ row percents in the shaded cells), the coding of variables is usually consistent with the coding of other variables.

[3]Code 49 is technically incorrect at this writing, because since at least 1992, it has been defined to indicate an ambulatory surgery center. We mention it here because at the time some UPIN records were added, it was a code used to indicate "miscellaneous" specialty.

6.0 CONSISTENCY ACROSS PRACTICE SETTINGS

In this chapter, we investigate the degree to which there is conflicting or missing information among the records of a single provider. In our examination, we performed three types of analyses on selected data elements to determine the number of providers (with multiple active practice settings) that: (1) had only missing values, (2) had a mixture of feasible and missing values, and (3) had inconsistent feasible values. Conflicts across multiple active practice settings, the subject of our analysis, could occur among the settings of a single carrier or among the settings of multiple carriers. A single UPIN's records could contain both types of conflicts. The extent to which values are missing across all active practice settings reveals which data elements were the most difficult to obtain. An examination of missing values in conjunction with feasible values tells us whether it is possible to migrate values from other practice settings.

Our investigation of consistency across practice settings reveals the extent to which provider information is identical across active practice setting records. Inconsistencies have implications both for the operational aspects of the Registry and for research using Registry data. For example, date of birth values should be identical across multiple active practice settings. Conflicting birth dates call into question the integrity of the UPIN as a *unique* identifier. In research, graduation year is often used as a proxy for provider experience. When individual Medicare claims are linked by UPIN to Registry data, inconsistencies between graduation year create problems for researchers since they will not know which is the actual value.

6.1 Methodology

6.1.1 Selection of Records and Data Elements for A

All providers analyzed in this chapter had at least two active practice ... records in the Registry file of July 1993. We considered only *active* practice settings when comparing information across practice settings because they are usually generating claims, thus it is more important that their provider information be correct. Out of the 693,355 UPINs, 427,126 had at least two active practice settings. Comparing multiple practice settings allows us to determine the extent to which UPIN data elements are inconsistent across practice settings.

Using this sample, we chose to include in our analysis data elements with high MEDU rates (see Tables 4-1 and 4-2 in Chapter 4), excluding state license number. In addition, the three sanction data elements and date of death were included because of the possibility that some carriers would not be updating their records in a timely fashion. Provider credentials and physician status were also examined because they are often used to classify providers in research. We examine only the setting data elements (see Chapter 7 for a discussion of inconsistencies between header and setting data elements). Our analysis focuses solely on UPINs with all missing values for a given data element, all feasible values, or combinations of missing and feasible values across carriers and across multiple active practice settings.

6.1.2 Definitions of Numerators and Denominators

In examining UPINs with all missing values for a given data element, we calculated the percent of UPINs with missing values across all of their active practice settings. The denominator is the total number of UPINs with at least two active practice settings. In most cases, missing values include blanks in the field as well as instructed codes for missing values (see Exhibit 3-3 for definitions of missing values for each data element). Whereas a blank does not contain any characters or numeric values, an instructed code for missing values might be

zeroes or nines. (See the discussion below of sanction data elements and date of death for special cases.)

Feasible values are those which appear to be valid; in this analysis, dubious and erroneous values were eliminated. Our goal in dropping the dubious and erroneous values was to measure the extent of problems and possible clean-up efforts beyond values which are simply incorrect. Our investigation of feasible values focused on two areas: inconsistency among feasible values and feasible values in conjunction with missing values.

In our analysis of feasible values, we checked for any inconsistency among practice setting values. For a given data element, if one datum conflicted with any other practice setting value in our sample, then that UPIN was labeled as containing an inconsistency. Because there is only one true value for each data element, consistency checks reveal whether a given UPIN contains an error in at least one of the practice settings. Note that for several data elements, the base of feasible values is much larger for all feasible values than for only those greater than zero/blank. Additionally, the base for feasible date values is much larger when only feasible year values are required than when one requires full date feasibility.

Finally, we pool UPINs with all missing values, UPINs with feasible values, and UPINs having a combination of missing and feasible values among their active practice settings. From this denominator, we look for UPINs having a mixture of missing values and feasible values among their active practice settings. This information tells us whether it is possible to use values from practice settings with feasible values to fill in those practice settings which are missing a data element.

Note that this correction process would require consistency among feasible values across practice settings. If a UPIN contains missing values and inconsistent feasible values for a data element, then it is not immediately clear which feasible value to use to fill in missing values. Our analysis does not show the percent of UPINs that have consistent feasible values and missing values across practice settings, but rather the percent of UPINs with feasible

(consistent or inconsistent) and missing values out of a total denominator of UPINs with all missing values, all feasible values, and combinations of both.

6.1.3 Special Procedures for Dates and Sanction Variables

Several data elements required multiple levels of analysis. In order to pinpoint specific data problems, we divided the date data elements birth year, death year, and sanction date into year only and full date categories. This allows us to investigate whether inconsistencies are present in the entire date or in only a portion of the date. Carriers were instructed to zero-fill any unknown portions of date data elements. Therefore, a comparison of year only and full date categories also reveals the extent to which month and day values have been zero-filled.

In cases where a data element is not applicable (e.g., date of death, sanction date, sanction code, sanction length) for a given provider, carriers were instructed to code zeroes for this value (a blank in the case of sanction code). In these cases, one cannot distinguish between non-events, partially-known events, and known events. As illustration, a sanction date of 0000 could represent a provider who is not sanctioned or a sanctioned provider with an unknown sanction date. Moreover, if the year component of the field remains as two digits, then in the future it could represent a provider sanctioned in the year 2000 but the month is unknown.

Due to the dual function of zero for date of death and the sanction data elements, zero-filled values were considered feasible. In our analysis, we stratified these data elements into: (1) feasible and (2) feasible non-blank/non-zero-filled categories. *All feasible values* includes feasible values and blank- or zero-filled values (indicating either a non-event or an unknown event). The category *feasible values greater than zero* (or not blank/not zero-filled) captures non-zero (non-blank) values that are feasible and indicate that the data element is applicable. While only the latter category is a suitable base for capturing recording inconsistencies among

known events, it is not a suitable base for capturing inconsistencies among all events, for it excludes events which are missing information.

The dual function of zero-filled values to indicate both "not applicable" and "missing" also caused problems when calculating the percentage of UPINs with missing values for all practice settings. That is, for date of death and the sanction data elements, nearly all values were equal to zero, thus giving the misleading impression that this information was missing for most UPINs. In light of this, we indicated that this calculation is not applicable for these data elements.

6.1.4 Treatment of Railroad Retirement Board Carriers

We created two tables which examine inconsistencies across practice settings, one with all carriers and one which excludes Railroad Retirement Board (RRB) carriers. The RRB carriers process claims from all states. This creates unique problems in updating provider information. Most other carriers have a working relationship with providers in their state which they can employ when attempting to obtain information. Moreover, if the provider is uncooperative, a carrier can consult local and state sources to update provider information in practice settings across its state. In contrast, the RRB receives claims from practice settings nation-wide and usually does not have the same working relationship with the providers they service. This makes it difficult to obtain the requisite information from providers. As a consequence, the RRB carriers are more likely than other carriers to have missing and/or erroneous values for a given data element. A table excluding practice settings submitted by the RRB carriers allows us to determine the extent to which inconsistencies were due to data collection difficulties confronting the RRB carriers.

6.2 Empirical Results

Tables 6-1 and 6-2 show the results of comparisons for nine data elements on the UPIN Registry. Table 6-1 includes all carriers, and Table 6-2 excludes RRB carriers. The percentage of inconsistencies by carrier[1] is shown in Appendix C tables. The remainder of the chapter will present a data element-by-data element analysis of Tables 6-1 and 6-2, followed by a brief discussion of the carriers containing the highest numbers of inconsistencies.

Because blanks/zeroes serve a dual function when used for certain data elements, calculating the percent of UPINs with all missing values (i.e., all blanks/zeroes) is not appropriate for our purposes. Therefore, in Tables 6-1 and 6-2, the date of death and all three sanction data elements show N/A (not applicable) for the *Percent of UPINs with All Missing Values*.

For date of birth, school code, and graduation year, the counts of UPINs in column two in Table 6-1 (and Table 6-2) are substantially lower than in column four. The reason is that the total UPINs with feasible values (column two) excludes UPINs with all non-feasible values (missing as well as erroneous or dubious values). Unlike the other data elements in Tables 6-1 and 6-2, the missing value rates are high for date of birth, school code, and graduation year. Column four, however, only excludes UPINs whose values consist only of erroneous values or dubious values.

It is important to note that not all inconsistencies are indicative of carrier errors. That is, the data analyzed are from one point in time whereas the maintenance of Registry data occurs over time. For instance, when a provider's record is updated and multiple carriers service the provider, differences in carrier timeliness in updating individual practice settings can lead, at least temporarily, to inconsistent values across practice settings.

---

[1] Appendix C Tables report within-carrier rates only, so that counts will not add up to counts derivable from Table 6-1.

TABLE 6-1

COMPARISONS OF VALUES ACROSS A PROVIDER'S MULTIPLE ACTIVE PRACTICE SETTINGS
(INCLUDING RRB CARRIERS)

Setting Variable	Percent of UPINs with All Missing Values*	FEASIBLE VALUES		Total Number of UPINs with either Feasible or Missing Values	Percent of UPINs with Feasible and Missing Values
		Total Number of UPINs	Percent Inconsistent		
Date of Birth					
Year only	N/A %	347,096	3.95 %	426,459	28.36 %
Full date	6.04	336,197	7.02	415,530	28.06
Date of Death (all feasible values)					
Year only	N/A	427,072	0.02	427,072	0.0007
Date of Death (feasible values > 0)					
Year only	N/A	30	0.00	N/A	N/A
Full date	N/A	30	3.33	N/A	N/A
Medical School	6.96	337,713	11.23	424,575	29.84
Graduation Year	8.65	319,616	10.38	424,065	33.15
Sanction Date (all feasible values)					
Year only	N/A	427,088	0.06	427,088	0.06
Sanction Date (feasible values > 0)					
Year only	N/A	248	4.03	N/A	N/A
Full date	N/A	245	10.61	N/A	N/A
Sanction Code (all feasible values)	N/A	427,126	0.31	427,126	0.30
Sanction Code (feasible, not blank-filled)	N/A	444	3.83	N/A	N/A
Sanction Length (all feasible values)	N/A	427,101	0.07	427,101	0.00
Sanction Length (feasible values > 0)	N/A	255	2.35	N/A	N/A
Provider Status	0.00	427,125	0.14	427,125	0.00
Provider Credentials	0.00	427,126	3.19	427,126	0.00

* Percent of UPINs with all missing values is out of 427,126 UPINs with at least 2 active practice settings.
N/A -- not applicable
Comparisons were made without respect to carrier
For each data element, "inconsistent" UPINs contained at least one value that was different across practice settings.

SOURCE: Physician Registry, July 1993.

TABLE 6-2

COMPARISONS OF VALUES ACROSS A PROVIDER'S MULTIPLE ACTIVE PRACTICE SETTINGS
(EXCLUDING RRB CARRIERS)

Setting Variable	Percent of UPINs with All Missing Values*	FEASIBLE VALUES		Total Number of UPINs with either Feasible or Missing Values	Percent of UPINs with Feasible and Missing Values
		Total Number of UPINs	Percent Inconsistent		
Date of Birth					
Year only	N/A %	271,635	1.90 %	322,945	17.26 %
Full date	6.98	262,667	3.95	313,519	16.94
Date of Death (all feasible values)					
Year only	N/A	323,390	0.01	323,390	0.0006
Date of Death (feasible values > 0)					
Year only	N/A	30	0.00	N/A	N/A
Full date	N/A	30	3.33	N/A	N/A
Medical School	8.01	264,526	5.10	321,376	18.95
Graduation Year	10.36	250,303	4.88	322,024	22.27
Sanction Date (all feasible values)					
Year only	N/A	323,406	0.04	323,406	0.04
Sanction Date (feasible values > 0)					
Year only	N/A	230	3.48	N/A	N/A
Full date	N/A	227	7.93	N/A	N/A
Sanction Code (all feasible values)	N/A	323,423	0.30	323,423	0.30
Sanction Code (feasible, not blank-filled)	N/A	409	2.44	N/A	N/A
Sanction Length (all feasible values)	N/A	323,422	0.04	323,422	0.00
Sanction Length (feasible values > 0)	N/A	240	0.83	N/A	N/A
Provider Status	0.00	323,423	0.06	323,423	0.00
Provider Credentials	0.00	323,423	0.87	323,423	0.00

* Percent of UPINs with all missing values is out of 323,423 UPINs with at least 2 actve practice settings.
N/A -- not applicable
Comparisons were made without respect to carrier.
For each data element, "inconsistent" UPINs contained at least one value that was different across practice settings

SOURCE: Physician Registry, July 1993.

This is especially the case for date of death and the sanction data elements. Unlike other data elements which should be present when a new practice setting is added, date of death and the sanction data elements denote events that need to be discovered and then verified – actions that carriers perform over time. For date of death, it is often the case that one carrier is the first to discover that a provider is deceased. If other practice settings are maintained by other carriers, the other carriers need to verify the date of death before changing the information on their practice settings for the provider. Thus, a lag between the initial discovery of a provider's death by one carrier and verification by other carriers is almost inevitable. Since carriers are simultaneously notified about *new* sanctions by HCFA, only differences in timeliness in identifying the provider's UPIN and submitting updates to the Registry *should* account for temporary inconsistencies. (Note that the preceding discussion does not apply to events that occurred before the establishment of the Registry.)

6.2.1 Date of Birth (DOB)

DOB is a six-digit numeric data element to be entered as MMDDYY. Thus, feasible values for month and day would include 0101 through 0131 for January birthdates, 0201 through 0229 for February, and so on. Feasible values for year of birth are 01 through 69 and 00. We chose 69 as the birth year cutoff assuming that an active 1993 provider could not be younger than 24. As a required field, DOB should contain no missing or unknown values, but '000000' is the instructed missing value.[2] All other characters would be errors and therefore infeasible. In addition, if only the partial date is available, usually the year, carriers were instructed to zero-fill the month and day portions.

[2]For date of birth our definition of missing values included blanks and values of all zeroes or all nines.

For date of birth, approximately six percent of UPINs (or 25,798) had all missing values for date of birth, full date.[3] As one would expect, the percent of UPINs with inconsistent feasible values was highest for the full date category at seven percent, or 23,601 total UPINs as opposed to 13,710 UPINs with at least one inconsistency in year values. For full date of birth, not only should the year value be identical across practice settings, but also the month and day values.

From the large percent of UPINs containing both missing and feasible values, we conclude that 120, 944 UPINs have practice settings for which they might obtain feasible values for birth year from another practice setting. Approximately 116,598 UPINs might obtain values for full date of birth in this same manner.

The total of UPINs with all missing values decreased from 25,798 to 22,575 when RRB carriers were excluded from the sample (Table 6-2). On a relative basis, missing values for date of birth across all of a provider's active practice settings were less apparent among RRB practice settings than among the remainder of the practice settings. On the other hand, the proportion of UPINs with inconsistent feasible values fell considerably when RRB carriers were excluded in both year only and full date categories. UPINs with both feasible and missing values were also significantly less without RRB carriers, indicating that these problems appeared more often among RRB practice settings.

Among carriers, Aetna Life and Casualty (Georgia), had the highest percentage of all missing values for the full date of birth at 37 percent, followed by C & S Administrative Services of Massachusetts with 32.5 percent (Table C-1). Kentucky BC/BS contained the highest percentage of feasible but inconsistent values with 28 percent. All other carriers had within-carrier inconsistency rates of nine percent or lower.

[3]We did not perform this calculation for year only.

6.2.2 Date of Death

DOD is a six-digit numeric data element entered as MMDDYY. As with DOB, the feasible month and day values should include valid combinations of months 1-12 and the corresponding appropriate number of days, varying from 29 to 31 depending on the month. The carriers were instructed to zero-fill this data element if the physician is not deceased. Feasible year values for date of death include 86-93, 99, and 00. Zero indicates a living provider in addition to cases where the carrier does not know the date of death.[4] (For further discussion of the definitions regarding feasible years of death, see Chapter 3.)

The dual function of zero-filled values to indicate both living providers and missing dates of death prevented an accurate assessment of truly missing values across all practice settings.[5] That is, because almost all providers were living, nearly all dates of death were equal to zero, thus giving the misleading impression that date of death was missing for most UPINs. In light of this, we indicated that this calculation is not applicable for this data element. Similarly, among non-zero feasible values, we did not calculate the percent of UPINs with feasible and missing values. (Non-zero feasible value specifications would eliminate zero-filled entries from our sample.)

For *all feasible values*, we chose to examine consistency among year values only. Full date specifications would eliminate zero-filled entries, which are feasible according to our definitions, from our sample. Among year values, .02% of 427,072 feasible values were inconsistent. Among all feasible values, inconsistencies in death year may be attributable to differences in null values (i.e., both "00" and "99" are present), non-identical feasible values, discrepancies such that some practice settings have zero-filled values while others have feasible values, or combinations of these inconsistencies. From the 427,072 all missing, feasible, and

[4]We allowed dates filled with nines to be the equivalent of zero-filled dates.

[5]For date of death, only blanks were considered missing values.

mixture of feasible and missing values, .0007% (3) were inconsistent, indicating a possibility for gap-filling in only three cases.

In the *feasible values greater than zero* category, the number of UPINs with feasible values drops from almost the entire sample in *year only* to 30 for *full date*. This is a result of excluding the high number of zero-filled values which do not meet the criteria for feasible full date. In the feasible but inconsistent category, one UPIN had at least one inconsistency between the month and day values across practice settings. When RRB carriers are excluded from the sample, the total number of UPINs with inconsistent feasible values drops from 85 to 32 and the inconsistency percent drops by half. Among non-zero values, one UPIN contains an inconsistency among feasible full date values for date of death.

Only 30 feasible dates of death appeared in the sample because we were examining practice settings where provider DRIP status was active or missing.[6] When a provider dies, the DRIP status soon becomes deactive; thus, most such cases would not have shown up in our sample. The small number which appeared in our point-in-time sample might be due to some carriers' practice of keeping a deceased provider's DRIP status active for a short period following his/her death. This allows for outstanding claims to be processed by the carrier. Due to the low number of UPINs with non-zero feasible dates of death, we did not conduct a carrier-by-carrier analysis of all-missing and inconsistency rates.

6.2.3 Medical School

Medical school is a five-character alpha-numeric code which carriers were instructed to "blank fill" if the data were unavailable. Feasible values for medical school code included valid codes as well as catch-all codes used when only the type of medical school, but not the actual

[6]DRIP status is the type and/or activity of the practice setting: active practice setting, de-activated practice setting, and missing. In the Registry's documentation, DRIP status is named MPIER-RES-INTERN-CODE and has the following coding: D - De-activated [practice setting], R - Resident, I - Intern, P - [active] Practice [setting]. (See the discussion in Chapter 3.)

school, was known. All codes which were not classified as missing[7] or valid fell into the erroneous, or infeasible, category.

Medical school and graduation year were not required data elements for the initial load. As noted in Chapter 2, carriers sometimes find it difficult to obtain such information from providers unless a new practice setting is being added. Nearly seven percent of UPINs (29,728) had missing values for medical school code across all active practice settings, and 11.23 percent of UPINs with feasible values (37,925) had inconsistent data across practice settings. Though 29.84 percent (126,693) of values for medical school were feasible in some practice settings and missing in others, the high number of inconsistent feasible values may make corrections difficult. Feasible values cannot be used to fill in other practice settings until the correct feasible value is determined.

The percent of UPINs with all missing values increased from seven to eight percent when RRB practice settings were excluded from the sample, revealing that all missing values were a somewhat smaller problem among RRB carriers. UPINs with inconsistent feasible values fell by half from 11.23 percent (37,925) to 5.10 percent (13,491), indicating that these inconsistencies were largely due to RRB practice settings. UPINs with both feasible and missing values among their records were also highly prevalent among all carriers at 30 percent (126,693), though this percent dropped to 19 percent (60,901) when RRB carriers were excluded.

Within-carrier analyses show that Aetna Life and Casualty (Georgia) and South Carolina BC/BS had the highest percentages of missing values across all multiple active practice settings, at 37 percent and 32 percent, respectively. The RRB (Utah) and North Dakota BC/BS (Wyoming), had the highest percentage of inconsistencies between feasible values at 15 percent (Table C-2).

[7]Missing values were either blank or the instructed codes of 00000 or 00001.

6.2.4 Graduation Year

Graduation year is a two-digit numeric data element with feasible values ranging from 40 to 93. The cutoff of 1940 is based on the assumption that we should not have any providers over 77 years old still practicing in 1993. (Recall from the date of birth discussion that we assume the youngest medical school graduate is 24.) Missing values are 00 and 99, and infeasible values are all other values.

Like medical school code, there were many inconsistent and missing values for graduation year. Approximately nine percent of UPINs (36,946) had missing values for all practice settings for graduation year. Ten percent with feasible values (33,176) had data which were inconsistent. Although there was a large number of cases (140,578) with feasible values in some practice settings and missing values in others, the high number of inconsistent feasible values may make corrections difficult.

When RRB carriers were excluded from the sample, the percent of UPINs with all missing values increased to 10.36 (33,507), attesting that missing values across all practice settings were less of a problem within the Railroad Retirement Board practice settings. As with date of birth and medical school, inconsistent as well as feasible and missing values appeared more often across practice settings in RRB carriers.

Among carriers, Wisconsin Physician Service, TOLIC, and Aetna Life and Casualty (Georgia) had the highest percents of all missing values: 50.5 percent, 39 percent, and 37 percent, respectively (Table C-3). Indiana BC/BS had the highest inconsistency rates at 16 percent, followed by the RRB (Utah) with 14 percent.

6.2.5 Sanction Date

Sanction date is a four-digit numeric data element entered as MMYY on an if-available basis. There were no specific values assigned for missing or unknown, but carriers were

instructed to zero-fill dates. Thus, one cannot distinguish "not sanctioned" from a sanction date that was missing. We indicated that the percent with all missing values was not applicable. Due to the fact that full date specifications did not define as "feasible" zero-filled values for month and day, we only examined sanction date feasibility among year values for the *all feasible values* category. Furthermore, for *feasible values greater than zero,* we indicated that the percent of UPINs with feasible and missing values was not applicable. (This is because zero-filled values, otherwise feasible, would not be included in our counts, making the calculation inappropriate.)

For *all feasible values*, only .06% of the providers had inconsistent year values, a total number of 256 UPINs. The same number of UPINs appeared again among UPINs with both feasible and missing values across their practice settings, suggesting that these UPINs had an information conflict in that one or more setting records indicated a sanction occurred, while other records indicated no sanction occurred. For the *feasible values greater than zero* category, inconsistencies dropped to 10 for year only and 26 for full date.

When RRB practice settings were excluded, UPINs with inconsistencies among all feasible values fell from .06 (256) to .04 percent (129) (Table 6-2), indicating that this type of inconsistency was more common among RRB carriers. Among non-zero feasible values, inconsistency rates also dropped to 3.48 percent (8) for year only and 7.93 percent (18) for full date.

Table C-4 shows that the inconsistencies for full date, feasible values greater than zero were due to only a few carriers. Blue Shield of Western New York had one case of feasible values across practice settings, though these values were inconsistent, giving them an inconsistency percent of 100. Aetna Life and Casualty (Arizona) had two cases of feasible sanction dates across practice settings. One case contained inconsistent feasible values, giving them a 50 percent inconsistency rate.

6.2.6 Sanction Code

Sanction code is a one-digit alpha-numeric code allowing codes A through R and U (for unknown) as feasible values. However, carriers were instructed to blank fill values if sanction code was not applicable[8]. As with sanction date, we indicated that the percent of UPINs with all missing values is not applicable. We also indicated that the percent of UPINs with feasible and missing values was not applicable, for blank is excluded as an acceptable feasible value in the *feasible, not blank-filled* definition.

All of the UPINs in our sample had feasible values for sanction code, as indicated by the 427,126 UPINs included in the base. When blanks were excluded, the base dropped to 444. Inconsistent UPINs dropped from 1,324 to 17. This large drop in inconsistencies suggests that many inconsistencies are attributable to cases of a recorded sanction code in at least one practice setting and blank codes in other. Note that the *percent* of inconsistencies increases from all feasible values to non-blank feasible values. A possible explanation for this increased percent is that non-sanctioned providers with blank values across all practice settings augmented the proportion of consistent values in the first instance, but not in the second. The percent of UPINs with feasible and missing values out of the total number with either feasible or missing values was .30 percent of the entire sample, or 1,281.

When RRB carriers were excluded, inconsistencies among all feasible values fell from 1,324 to 970, but the percent of inconsistencies among the sample remained relatively constant at .31 and .30, respectively. Non-blank feasible but inconsistent values dropped from 3.83 to 2.44 percent. The percent of UPINs with missing values in some practice settings and feasible values in others remained the same, indicating that this problem existed in the same proportions among RRB practice settings as it did among practice settings for all carriers.

Within their own records, only three carriers contained inconsistent feasible values for sanction code (data not shown). The RRB (Utah) contained two cases of non-blank feasible

[8]Blanks and zero-filled values were considered feasible.

values greater than zero across practice settings, of which one contained an inconsistency. Texas BC/BS had five cases of non-blank feasible sanction codes, of which one contained inconsistent values. The final carrier containing this same type of inconsistency was Florida BC/BS at 2.5 percent out of 40 cases with non-blank feasible sanction codes.

6.2.7 Sanction Length

Sanction length is a two-digit numeric data element representing the length of a provider's sanction. Sanctions are normally given in five year increments, though sanction lengths of 1-14 were also viable. Thus, we considered sanction lengths of 1-15, 20, 25, 30, 35, 40, 45, and 50 feasible values. As with date of death and the other sanction data elements, missing or zero-filled values served a dual function, therefore they were considered feasible. We did not calculate all missing values for sanction length due to the large number of UPINs with zero-filled entries for non-sanctioned providers. Nor did we calculate the percent of UPINs with feasible and missing values in the *feasible values greater than zero* category.

The number of UPINs with feasible sanction lengths dropped from 427,101 to 255 when zero-filled values were excluded. Whereas .07 percent of UPINs (299) with feasible sanction lengths had inconsistencies across practice settings, approximately two percent of UPINs (6) with non-zero feasible values had inconsistencies for sanction length among their practice settings. Like sanction code, this drop in inconsistencies among feasible values suggests that carriers have failed to process all relevant practice settings when updating provider information. These inconsistency rates decreased to .04 (129) and .83 percent (2), respectively, when RRB practice settings were excluded. In both Table 6-1 and 6-2, no UPINs had feasible values for sanction length in some practice settings and missing values in others.

Two carriers contained within-carrier inconsistencies between multiple active practice setting values for sanction length, feasible values greater than zero (data not shown). The

Railroad Retirement Board (Utah) contained two cases of feasible values greater than zero for sanction length, and one case contained inconsistencies. Texas BC/BS contained inconsistencies in one of its five cases of feasible sanction lengths.

6.2.8 Provider Status

The provider status indicator is a single numeric digit with allowable values of 1 for MD/DOs and 2 for all other limited license practitioners. There are no assigned values for missing or unknown. Values other than one or two were considered erroneous.

Provider status was fairly consistent across practice settings. No UPINs were missing physician status across all active practice settings. The percent of inconsistent feasible values, .14 (598 UPINs), did not include any cases where some practice settings had feasible values while others were missing. When RRB carriers were dropped, the inconsistency rate decreased to .06 percent (194 UPINs), revealing that the problem was largely associated with RRB practice settings.

Table C-5 shows inconsistency rates in provider status within the carriers. Due to the fact that almost no UPINs were missing values for this data element, we calculated inconsistencies among feasible values only. Aetna Life and Casualty (Alaska) had the highest percent of inconsistencies at .28 percent, followed by the two RRB carriers at .27 percent (Utah) and .25 percent (Georgia).

6.2.9 Provider Credentials

Provider credentials is a three-digit alpha-numeric code with allowed values of MD, DO, DDM, DDS, DPM, OD, and CH. There are no assigned codes for missing or unknown, but a blank is the implicit missing value.

The entire group of multiple active practice settings contained feasible values for provider credentials. Not only did no UPINs contain missing values across all practice settings, but also no UPINs had missing values for provider credentials in any practice setting. The percent inconsistent was 3.19 (13,625 UPINs), dropping to .87 percent (2,814 UPINs) when RRB practice settings were excluded.

Among carriers, we did not calculate rates of missing values because no UPINs contained missing values for provider credentials. Colorado BC/BS had the highest rate of inconsistencies across practice settings at 6.09 percent, followed by Kansas BS (Missouri) with 3.69 percent and Iowa BS with 2.5 percent (Table C-6). The remainder had inconsistency rates of two percent or lower.

6.3 Conclusions

Reasons for the observed inconsistencies can be classified into two categories: (1) time lags in obtaining information which only pertains to some providers and (2) difficulty obtaining and disseminating information which is relevant for all providers. Whereas death year inconsistencies probably resulted from time lags in changing provider information as it became available, inconsistencies among sanction data and data elements relevant to every provider (e.g. birth year) revealed more fundamental difficulties encountered by carriers.

One difficulty in maintaining sanctions data is recognizing a sanctioned provider as one for which the carriers maintain Registry records; the sanctions notices issued by the Office of the Inspector General do not yet include the UPIN, which would greatly ease identification. Another possibility is that during the implementation phase (1989), many carriers had only partially correct sanction data in their files for providers sanctioned prior to 1989.

Date of birth (DOB) had a high rate of inconsistent feasible values among multiple active practice settings. For full date, inconsistencies among feasible values were a problem for

23, 601 UPINs. This is especially problematic because DOB is a critical data element to provider identification, especially in assuring that each provider only receives one UPIN. Similarly, inconsistent or missing dates of birth might result in more than one UPIN being assigned to a single provider (see Chapter 8).

Graduation year and medical school code had the highest rates of inconsistent values across feasible values of multiple active practice settings. In both instances, over 30,000 UPINs were affected. Because these data elements are used by the Registry to validate the physician's credentials with the American Medical Association before a UPIN is assigned, they are very important elements. Also, they could be used to measure provider experience and quality of education. Thus, determining the correct values for these data elements could significantly aid future research with Registry data.

Note that in cases of inconsistent feasible values among multiple active practice settings, consulting the header for the correct information would not be the appropriate course for amending the problem. Header information is arbitrarily constructed from practice setting records. Correct values must be obtained from the provider.

For birth year, medical school, and graduation year, nearly one-third of UPINs contained missing values in some practice settings and feasible values in others. Again, the large rate of non-identical feasible values for these data elements requires that the correct feasible one be determined before any values can be filled in across practice settings. Finally, these same three data elements contained significant rates of all missing values across practice settings. This might be indicative of the difficulties encountered by carriers when trying to obtain this information from providers.

7.0 REGISTRY DATA ELEMENTS COMMON TO THE RECORD HEADER AND PRACTICE SETTINGS

In this chapter, data elements common to the header section of each provider's Registry record and each practice setting segment are compared, in order to evaluate inconsistencies. The common data elements compared in this chapter are: date of birth, date of death, medical school, graduation year, sanction code, date when sanction took effect, and sanction length. In order to be compared, both the header and practice setting values had to be feasible (i.e., they had to meet our definitions for valid-looking data).

Header data elements are integral to Registry operations. In particular, the header date of birth and name fields are used to link new practice setting records to existing UPINs. If a match with an existing UPIN can not be made, a new UPIN might be issued. Discrepancies between feasible values of header and practice setting versions of data elements raise questions about the general accuracy of the Registry and could cause problems for users of the Physician Registry.

In documenting data for submission of a new setting record, for instance, carrier personnel are supposed to develop each data element on their own. However, carrier staff might consult Registry records as a starting point. If a feasible but inaccurate value is found (be it from the header or setting) it could mislead staff and ultimately lead to unnecessary carrier costs. It could also lead to additional errors in the Registry file. In general, inconsistencies among data elements weaken the confidence users have in Registry data. If too many users express complaints about the quality of Registry data, the usefulness of the database could be called into question.

Researchers have even fewer resources than the Registry or the carriers to resolve discrepancies between header and practice setting values. Even the best rules for choosing

which value to use will result in some errors. Such data shortcomings lead to weaker research findings.

7.1 Background and Methodology

7.1.1 Background

The file periodically delivered by TOLIC to HCFA has variable-length records with a header section followed by as many practice setting segments as a provider has practice settings (see Chapter 3). Since a provider may be serviced by multiple carriers, the values of header variables may come from more than one carrier. Except for name fields, when a new setting value for a data element is sent to the Registry, TOLIC updates the corresponding header value if the new setting information is feasible, regardless of whether it is consistent with the provider's other active practice settings.

Though TOLIC automatically updates header information from new feasible practice setting values, it does not automatically replace or validate existing practice setting information. Such processes make inconsistency rates between header and setting not only dependent on the accuracy of existing practice setting information, but also highly sensitive to the accuracy of incoming practice setting information.

7.1.2 Methodology

As in Chapter 6, de-activated practice settings were not included in the analysis. Comparisons were limited to data elements which were feasible in both the header and practice setting segments. For data elements that are not applicable to every provider (date of death and the sanction variables), we stratified our analysis into all feasible values and feasible values greater than zero. For feasible values greater than zero, both header and setting values which

were not greater than zero were excluded from our sample. (For further discussion of analytical problems regarding date of death and sanction date variables, please see Chapter 6.)

Inconsistencies within individual carriers are shown in the Appendix D tables. When examining these rates, it is important to keep in mind that they are dependent on the accuracy of incoming provider practice setting information. Given TOLIC's procedure of updating header information, a carrier that submits a new or updated practice setting to the Registry with feasible but incorrect provider information will affect the inconsistency rates between header and setting data elements for all carriers. Carriers that are able to update their practice settings more frequently may also be affecting inconsistency rates between header and setting values for other carriers.

7.2 Results

This section discusses, in turn, each of the data elements for which comparisons were performed. Each discussion of a data element presents the overall inconsistency rate and then indicates the carriers that had the highest percentages of inconsistencies. For the same reasons given in Chapter 6, two tables showing inconsistency rates are presented, one that includes all carriers and the other that excludes the two Railroad Retirement Board (RRB) carriers. (For further discussion of difficulties confronting RRB carriers, see Chapters 2 and 6.) Detailed information on inconsistencies by carrier appears in Appendix D. Note that the carrier identification in Appendix D is from the carrier on the setting involved in the comparison. There is no carrier data element on the header.

7.2.1 Date of Birth

To examine inconsistencies between header and practice setting date of birth values, we analyzed both the year portion of the date and the full date. (For further discussion of data

element format and definitions of feasible values, please see Section 6.2.1 in Chapter 6.) Among a denominator of all feasible values (1,408,935), 1.64 percent of the year comparisons (23,107 cases) were inconsistent between header and setting version (Table 7-1). Our full date requirements were more stringent, thus dropping the total cases of feasible values to 1,360,174. Of those, 2.92 percent were not identical across header and practice setting versions, representing 39,717 cases. Due to the large number of cases affected, we considered inconsistency to be a significant problem for the date of birth data element.

The percent of inconsistencies fell when RRB practice settings were excluded. Inconsistencies between feasible header and practice setting values fell to 1.40 percent for year values and 2.51 percent for full date values (Table 7-2). This reduction in inconsistencies reveals that a disproportionate share of the inconsistencies were between header and RRB practice setting versions of date of birth. However, there were still 16,430 cases of inconsistency among year values, and 28,325 cases of inconsistency among full date values.

Almost all carriers had some inconsistencies between header and setting versions of date of birth (Table D-1 shows full date results). Again, it is important to remember that individual carriers may not be responsible for inconsistencies between their header and practice settings. S.S.S. (Virgin Islands) had eight practice settings with feasible cases of date of birth. One practice setting had inconsistencies within the header value, thus giving S.S.S. a 12.50 percent inconsistency rate. They were followed by Utah BC/BS at 6.68 percent of 3,745 comparisons.

7.2.2 Date of Death

Capturing inconsistencies in date of death values across header and practice settings required some tailoring of our analytical methods. Due to the standard practice of zero-filling death year to indicate that a provider is living, we stratified the data element into two

TABLE 7-1

PERCENT INCONSISTENT ACROSS HEADER AND PRACTICE SETTING VERSIONS OF VARIABLES (INCLUDING RRB CARRIERS)*

Variable	FEASIBLE VALUES: Total Number of Cases	FEASIBLE VALUES: Percent Inconsistent
Date of Birth		
Year only	1,408,935	1.64 %
Full date	1,360,174	2.92
Date of Death (all feasible values)		
Year only	1,681,161	0.29
Date of Death (feasible values > 0)		
Year only	423	0.71
Full date	410	0.98
Medical School	1,368,362	5.01
Graduation Year	1,299,484	4.52
Sanction Date (all feasible values)		
Year only	1,681,861	0.05
Sanction Date (feasible values > 0)		
Year only	1,243	0.97
Full date	1,227	5.13
Sanction Code (all feasible values)	1,682,110	0.11
Sanction Code (feasible values > all blank)	1,614	1.73
Sanction Length (all feasible values)	1,681,938	0.05
Sanction Length (feasible values > 0)	1,253	0.72

* out of 1,682,110 total cases.

SOURCE: PHysician Registry, July 1993.

upin\tables\TAB7-1.XLS\sjb

TABLE 7-2

PERCENT INCONSISTENT ACROSS HEADER AND PRACTICE SETTING VERSIONS OF VARIABLES (EXCLUDING RRB CARRIERS)*

	FEASIBLE VALUES	
Variable	Total Number of Cases	Percent Inconsistent
Date of Birth		
Year only	1,173,570	1.40 %
Full date	1,128,500	2.51
Date of Death (all feasible values)		
Year only	1,360,480	0.19
Date of Death (feasible values > 0)		
Year only	419	0.72
Full date	406	0.99
Medical School	1,131,781	4.48
Graduation Year	1,070,064	4.07
Sanction Date (all feasible values)		
Year only	1,361,153	0.04
Sanction Date (feasible values > 0)		
Year only	1,178	1.02
Full date	1,162	4.99
Sanction Code (all feasible values)	1,361,311	0.12
Sanction Code (feasible values > all blank)	1,503	1.53
Sanction Length (all feasible values)	1,361,219	0.04
Sanction Length (feasible values > 0)	1,193	0.59

* out of 1,361,311 total cases.

SOURCE: Physician Registry, July 1993.

categories: all feasible values and feasible values greater than zero. Thus, *all feasible values* included feasible dates of death as well as zero-fill values for living providers. (For further discussion of data element format and definitions of feasible values, please see Section 6.2.2 in Chapter 6.)

In the *all feasible values* category, .29 percent of all cases (4,875) had different header and setting values for year of death. We did not compare the full date because full date specifications exclude "feasible" values of zeroes. The total number of feasible year values fell to 423 when non-zero values were excluded, indicating that 423 active setting-header pairs recorded a death.

Inconsistencies among all feasible values could be associated with a variety of events. A deceased provider may have some practice settings that recorded a death while others did not record the event (they zero-filled it), or the provider may have conflicting feasible values among practice settings. A living provider may have some practice settings filling in zeroes to indicate that this data element is not applicable while others filled in nines. When the sample is restricted to non-zero feasible values, all inconsistent cases associated with zeroes and nines were dropped. Only discrepancies between feasible values were captured.

There were slightly more feasible death years than full dates of death, 423 and 410, respectively. This is a result of some dates of death containing feasible year values but month and day values which are infeasible or zero-filled. In both cases, there were only a handful of discrepant results: .71 percent (3 cases) contained discrepancies between header and setting year values. For full date, .98 percent of 410 (4 cases) contained different values for header and setting.

When RRB carriers were excluded, the percent of inconsistencies between all feasible header and setting year values decreased by .1 percentage points to .19 percent. RRB carriers had more inconsistencies with the header date of death among their practice settings. Among

non-zero values, there were still three cases of inconsistent year values and four cases of inconsistent values across the full date of death.

Only two carriers (data not shown) had different header and setting values for date of death (full date, feasible values greater than zero). Illinois BC/BS had 32 cases of feasible dates of death, and 3 (9.38 percent) were inconsistent across header and practice setting versions. C & S Administrative Services of Massachusetts had 158 cases of feasible dates of death. Only in one case (.63 percent) was there a discrepancy.

7.2.3 Medical School

Among all the variables with header and setting components, medical school code had the highest rate of inconsistencies among feasible values. (For a discussion of data element format and definitions of feasible values, see Section 6.2.3 in Chapter 6.) Approximately five percent (68,554 cases) of all feasible values for medical school code were inconsistent. This was not unexpected given that this data element was not required during the implementation phase of the Registry. The inconsistency rate fell slightly to 4.48 percent (50,704) when RRB carriers were excluded from the sample. Difficulties confronting RRB carriers disproportionately increased the overall inconsistency rate.

All carriers experienced inconsistencies between header and practice setting medical school code values (Table D-2). North Dakota BC/BS carriers (Wyoming and North & South Dakota) had the highest inconsistency rates with 16.14 percent and 15.87 percent, respectively.

7.2.4 Graduation Year

Like medical school code, there was a high percentage of inconsistencies between header and practice setting values for graduation year. (For a discussion of data element format and definitions of feasible values, see Section 6.2.4 in Chapter 6.) Out of approximately

1.3 million cases of feasible values for graduation year, 4.52 percent (58,737 cases) were not identical between header and practice setting. This percent dropped to 4.07 percent (43,552 cases) out of 1.07 million when RRB carriers were excluded.

Though graduation year, like medical school, was not a required data element initially, present header and setting discrepancy rates suggest that updating procedures may be a significant problem in current Registry operations. Within carriers, although S.S.S (Virgin Islands) had the largest percentage (14.29 percent) of inconsistencies, this was among only seven cases of feasible values (Table D-3). Indiana BC/BS had 8.28 percent inconsistencies among their 20,541 cases of feasible values. All other carriers contained non-identical values between header and setting versions.

7.2.5 Sanction Date

Sanction variables, like date of death, often contained zero-fill values to indicate that this data element did not apply to the provider. Thus, zero was considered a feasible value[1]. We therefore stratified the variable into two categories: *all feasible values* and *feasible values greater than zero*, with the latter excluding all zero-filled values. (For further discussion of data element format and definitions of feasibility, see Section 6.2.5 in Chapter 6.) Because full date specifications excluded otherwise feasible values of zeroes and nines, we did not examine full date inconsistencies in the *all feasible values* category.

From a base of 1,681,861 feasible values, only .05 percent of all sanction date values (841 cases) were inconsistent between header and setting versions. The number of inconsistent cases fell considerably when feasible values were restricted to those greater than zero; for non-zero feasible values, about one percent of sanction year values were inconsistent, representing 12 cases of inconsistencies. Possible explanations for this large decrease in inconsistent cases include inconsistencies among records of non-sanctioned providers (some practice settings may

[1] Some carriers used 9s instead of zero-filling; for the purpose of this analysis, it was considered the equivalent of zero.

have nines and others zeroes) and that some carriers might have low priorities for updating sanction information, leaving some practice settings with feasible sanction dates and some with zero-filled sanction dates. The inconsistency rate increased to 5.13 percent when the base was restricted to feasible full date, increasing the total number of cases to 63.

All inconsistency rates decreased slightly when RRB carriers were excluded (Table 7-2). The one exception was in the *feasible values greater than zero* category, year only, where the percent inconsistent increased slightly by .05 percentage points when RRB carriers were dropped, indicating a better than average performance by RRB carriers in this instance.

Roughly one-third of the carriers had inconsistencies between header and setting values for sanction date (Table D-4). (Carriers with feasible sanction dates but no inconsistencies across header and practice settings were not shown in Table D-4.) Aetna Life and Casualty (Alaska) had one feasible sanction date between header and practice setting versions. These values were not identical, thus making their inconsistency rate 100 percent. Kansas BS (Missouri) had two cases of feasible values, and one had a discrepancy, thereby giving them a 50 percent inconsistency rate. Recall that these rates may be dependent upon the accuracy of practice setting submissions from other carriers and not necessarily the operations within an individual carrier.

7.2.6 Sanction Code

Because blanks were considered feasible, the entire sample contained feasible values for sanction code, though .11 percent (1,682 cases) were inconsistent. (For a discussion of data element format and definitions of feasible values, see Section 6.2.6 in Chapter 6.) However, in the *feasible values greater than all blank* category, 1.73 percent (32 cases) had inconsistencies. The large number of total inconsistent cases (a difference of 1,650) is partially attributable to some practice setting values being blank while the header value is non-blank and feasible. A

possible explanation is that some carriers might be failing to recognize that one of their providers is sanctioned while other carriers are putting more resources into maintaining sanction information. This could lead to updates of the header but persisting outdated information on other setting records for a single provider.

Inconsistencies among non-blank feasible values might include cases in which a sanction is known in one version and unknown (code = U) in the other as well as discrepancies among known codes between header and setting pairs. The percent of inconsistencies between feasible sanction code values was .01 percentage points lower when RRB carriers were included, though among non-blank feasible values, it was .2 percentage points higher.

Comparing sanction code to the other sanction variables, there were more feasible cases of sanction code between header and setting than there were for the sanction date and sanction length. Recall from Chapter 6 that feasible values for sanction code included blank values as well as codes of A - R and U for unknown. We did not have any responses that were outside of this feasible set, thus all values were considered feasible. Sanction date and sanction length, however, did contain infeasible responses, giving them a lower number of feasible cases from which to check for inconsistencies.

Only thirteen carriers had differences between feasible values for header and practice setting versions of sanction code (Table D-5). (Carriers with feasible values but no inconsistencies are not shown in Table D-5.) Sanctions are rare, so any cases of discrepant values resulted in high percents of inconsistency. For example, the highest percent of inconsistencies occurred in Blue Shield of Western New York. However, they only had 16 cases of feasible values for sanction code across header and practice setting versions. Prudential (North Carolina) had a similar scenario -- a 7.14 percent inconsistency rate within only 14 cases of feasible values. Some of these inconsistencies may be a result of one carrier's submission of an inaccurate but feasible sanction code for one of their updated or new practice settings.

7.2.7 Sanction Length

From a base of 1,681,938 feasible values, the percent of inconsistencies between header and practice setting segments was .05, representing 841 cases. (For a discussion of data element format and definitions of feasible values, see Section 6.2.6 in Chapter 6.) The base of feasible values fell to 1,253 for feasible values greater than zero, revealing that most providers were not sanctioned and therefore contained zero-filled values for sanction length. Among these cases, .72 percent (9 cases) were inconsistent. Inconsistency among sanction length values for sanctioned providers was not a significant problem for carriers.

Both inconsistency rates decreased when RRB carriers were excluded from the sample (Table 7-2). For all feasible values of sanction length, the inconsistency rate fell by .01 percentage points, and for *feasible values greater than zero*, the percent of different feasible values fell by .13 percentage points. Inconsistencies between header and setting values for sanction length, therefore, were more of a problem for RRB carriers than for others.

Only six carriers had feasible but inconsistent values for sanction length, feasible values greater than zero (Table D-6). (Carriers with feasible values but no inconsistencies are not shown.) Iowa BS had an inconsistency rate of 20 percent, meaning that one of their five cases of feasible values contained inconsistencies across header and practice setting versions (Table D-6). Wisconsin Physician Service had seven feasible cases, and their inconsistency rate was 14.29 percent. All other carriers had inconsistency rates of less than five percent. Note that some of these inconsistencies may stem from one carrier's submission of a feasible but inaccurate practice setting value, thereby changing the header information for this data element.

7.3 Summary

The results of the header and setting comparisons of data element values reveal that inconsistencies among feasible values are most serious for birth year, medical school, and graduation year. The higher rates of inconsistencies for medical school and graduation year might be explained by the fact that they were not required data elements initially. Thus carriers might not have taken as much care to ensure accuracy for these two data elements as for other data elements. Though the header versions of these elements are continuously updated as new feasible practice setting values are submitted, existing practice setting values, even if blank, are not automatically updated. Prior setting records often conflict with new header information, contributing to the high percent of inconsistency. Inconsistencies between feasible values for medical school occurred in 68,554 cases, and 58,737 cases had inconsistencies between feasible values for graduation year.

Although date of birth has been required since the beginning of Registry operations and is a key data element in identifying providers, inconsistencies between header and setting feasible values *full date* for this variable affect almost 40,000 cases. Given the importance of date of birth in issuing UPINs, identifying providers, and attaching new practice setting information to the correct provider, inconsistencies in date of birth information across practice settings could present serious complications to future Registry operations.

The automatic updating of header data elements leads to inconsistencies between header and practice setting values. It is performed even for practice settings submitted by Railroad Retirement Board (RRB) carriers. This practice seems inadvisable because the two RRB carriers do not have the same access as do other carriers to information on providers. Thus, RRB carriers might be disproportionately contributing to discrepancies between header and practice setting values.

8.0 MULTIPLE UPINS PER PROVIDER

Congress saw a need for Medicare to establish a unique identifier to support administrative functions such as uncovering fraud and abuse. Unique numbering also greatly facilitates research into topics including provider behavior and payment issues. To realize its potential as such a tool, the UPIN system should ensure that each provider has only one identification number.

Transamerica (TOLIC) has a set of procedures designed to prevent assignment of multiple UPINs to a provider. For incoming carrier records with a UPIN, the provider's last name and the first letter of the first name need to match. For carrier records with a UPIN that do not have the requisite name match and for carrier records submitted without a UPIN, TOLIC will try a computerized match against existing header information. The match is based on a key constructed out of the first three letters of the provider's last name, the first letter of the first name, a "soundex" value based on the remaining letters of the last name and, in some cases, a zip code.

If the "keys" match, then the computer algorithm checks practice setting information: the provider's last name, first name, middle initial, full date of birth, school code, graduation year, and state licensed in all must exactly match in order to attach the practice setting to an existing provider record. When checking the provider's existing practice settings, the algorithm checks one setting at a time. It does not borrow information that matched from previous settings that were checked.

If a match is not found, TOLIC then checks an American Medical Association (AMA) data extract. If a match still has not been found, TOLIC's system will flag the new practice setting for manual review. A TOLIC analyst will review existing UPIN records with the provider's last name providing the basis for the initial screening element. When reviewing records, the analyst has full access to both header and practice setting information.

In this chapter we present the results of our attempt to detect cases in which individual providers were issued two or more UPINs. Because variations in the detection algorithms produced a range of 2,617 to 25,308 multiple-UPIN cases, we assess the algorithms in terms of their "false-positive" rate. On a base of 587,305 UPINs that have active practice settings *and* feasible dates of birth (see section 8.1.2), the *net* potential number of cases suggests a 0.4 percent to 2.9 percent error rate. Also presented are characteristics of multiple-UPIN cases.

8.1 Methodology

The purpose of this chapter is to detect *cases* of multiple-UPIN providers. (In this chapter, the word *cases* refers to sets of two or more UPINs identified by the detection algorithms.) Since many of the data elements (e.g., initial letter of first name) that might be used to uniquely identify a provider are missing and/or have erroneous values, several variations of a detection algorithm are necessary. In general, it is desirable for the algorithms to detect as many true cases of multiple-UPIN providers as possible while minimizing the number of cases that, upon further examination, do not appear to be actual multiple-UPIN cases (we call these *false positives*). As will made clear by the discussions of data integrity and general conceptual issues, cases identified by the detection algorithms should be regarded as ***potential*** cases of multiple UPIN providers. To verify that an identified case indeed represents just one provider, at a minimum, visual inspection of the data elements present in each record in each case is needed. The necessity of visual inspection of potential cases implies that it is important, in order to hold administrative costs down, to avoid too many apparent false positives.

8.1.1 Data Elements

To detect cases of multiple UPINs for a provider, a "screener" variable (SCREENER) was created from each record. SCREENER resulted from concatenating parts of several variables to form one new variable. The last name of the provider was also used to calculate a check variable, LNAMESUM, discussed below. When two or more records have identical SCREENERs, a multiple-UPIN case is suspected. As the following discussion will show, several components of SCREENER, while used to guard against false positives, cannot altogether prevent the dropping of good suspects, leading to underdetection of potential cases at the same time that some false positives are eliminated.

In actual use, the cases uncovered by SCREENER mark just the first step in identifying cases of a multiple UPIN. At the carrier level, a setting record for each member of a suspected case identified by the SCREENER variable should then be compared visually with the records for other members, as described in the last paragraph (see also Section 8.1.5). Visual inspection should then show whether key elements of the record conflict or match, leading to a determination to either drop the case as a false positive or pursue it further as a suspected multiple-UPIN case. (We visually inspected a small sample of cases to evaluate the detection algorithm. The considerations we used in judging a case a false positive are discussed in section 8.1.4.)

SCREENER relied primarily on the date of birth and the last name of the provider. Variants of SCREENER also included the first letter of the provider's first name and the provider's credential. The *full* variant of SCREENER included, in order: date of birth, the first three letters of the last name, the first letter of the first name, LNAMESUM, and credential. LNAMESUM is the sum of numbers assigned to the consonants in the last name. Except for the credential, all variables were taken from the header section of a record. This was done to minimize problems due to practice setting variation.

Date of birth (DOB) was always included in SCREENER because of the large number of individuals with common last names such as Jones and Smith. The first three letters of the last name (L1_3) were also always part of SCREENER. Only the first three letters of the last name were used because of concern about spelling variations.

Many different names begin with the same three letters, so in some variants of screener LNAMESUM was used to avoid producing too many false positives. LNAMESUM was restricted to consonants because we believe that vowels are the most likely source of spelling variation.[1] (Therefore, LNAMESUM can fail to exclude some false positives, as when Ander*son* and Ander*sen* are tagged by the computer program as being identical.) In some circumstances LNAMESUM might overlook a good candidate for investigation. For instance, if an extra consonant was inadvertently added to a last name, as when an extra *L* is placed at the end of O'Neal (resulting in the spelling O'Neall instead of O'Neal) the screener values differ, and no multiple UPIN is detected.

The first initial of the first name (F1) was used in some variants of SCREENER because the number of providers with common last names is so high that even the inclusion of date of birth is not sufficient to distinguish between different providers. A difficulty with the first initial, of course, is due to the use of nicknames when the first letter of the nickname is not the same as the first letter of the proper first name (e.g., when *Bob* is used instead of *Robert*). All else being equal, the screener values would then differ, and no multiple UPIN would be detected.

The last element of SCREENER is provider credential (CREDENTIAL). Credential was included, again, because of the large number of individuals with the same first and last names as well as date of birth. Credential also has possible difficulties because about three percent of UPINs have inconsistent credential values (Table 6-1).

[1] A value of one was assigned to B, two to C, three to D, and so forth, with 21 assigned to Z.

8.1.2 Other Date of Birth Issues

Of the 693,355 UPINs, more than 64,000 had dubious, erroneous, or missing dates of birth with missing accounting for about 63,000 (Table 4-1). Additionally, in preliminary computer runs, ten potential cases of a multiple UPIN were found in which January 1, 1901, was indicated as the date of birth. (All ten cases were apparent false positives.) Since other data elements in the provider's record (e.g., graduation year) were not consistent with this birth date, we suspect that the date *January 1, 1901* was used by carriers when they did not know the provider's actual birth date and wanted to avoid TOLIC rejecting their submission. Therefore, only *feasible* dates of birth after January 1, 1901 were used in the algorithms to detect multiple UPIN cases.

Dropping provider records with infeasible dates and dates prior to January 2, 1901 probably results in fewer detected multiple UPIN cases. The loss of nearly 64,000 cases for checking due to date of birth restrictions make the DOB element of SCREENER less reliable than desired. Another problem with date of birth that can lead to fewer detected multiple UPIN cases is that the wrong date was entered (e.g., 11 instead of 12 for the month of birth). For instance, one carrier might have submitted 11 for the month instead of the 12 submitted by another carrier, which results in the detection algorithm failing because the dates of birth do not match.

8.1.3 Detection Strategy for Multiple-UPIN Cases

Since variation in the spelling of names is an important reason why the detection algorithms fail to detect multiple UPIN cases, variants of SCREENER were employed in separate computer runs. Date of birth and the first three letters of the last name were always used as part of SCREENER. Variations of SCREENER were produced by dropping, singly or in combination, the other components. Eight variations of SCREENER were eventually utilized.

Use of eight SCREENER variations results, then, in a range in the number of possible multiple-UPIN cases. When fewer components of SCREENER were utilized, we identified more potential multiple-UPIN cases; however, both the number and share of false positives increased as well. Although there are undetected multiple-UPIN cases and false positives, we believe the range in the number of possible cases is suggestive of the actual number of multiple-UPIN cases that would ultimately be confirmed upon local investigation.

8.1.4 Use of Practice Setting Information to Identify False Positives

To help assess whether or not a computer-identified multiple-UPIN case is a false positive, in addition to data from the two header portions of the records, information from the two practice settings was printed. Among the practice setting data elements printed were state license number, state licensed in, carrier code, credential, school code, graduation year, date of birth, business zip code, billing zip code, and all of the current name fields. These values were compared across the two practice settings, and where appropriate, to their header counterparts. In most instances, a discrepancy in one pair of *feasible* values (across carriers) results in marking the potential multiple-UPIN case as a *seemingly* false positive. There are instances, however, when it was necessary to take the available information in toto in order to make a decision. (An example is presented in section 8.2.1.)

8.1.5 Information from Active Practice Settings

To assess whether the potential multiple-UPIN cases had common attributes, characteristics of the cases were tabulated. The analyses (e.g., frequencies on the carriers involved in multiple-UPIN cases) rely partly on information (data) from the practice settings. Because it is often the case that there are multiple practice settings for individual UPINs, it was

necessary to use data from just one practice setting.[2] The first active practice setting added to the UPIN was chosen. To see why, note that multiple-UPIN cases arise when a carrier submits information to TOLIC and TOLIC is unable to find the provider in the Registry. That is, the provider already has an UPIN but the new information sent by the carrier is sufficiently different from that already in the Registry so as to cause the issuance of another UPIN to the provider.

Since header data for each Registry record relies, in the first instance, upon data supplied in the first submitted practice setting, it is likely then, that identification problems are the result of inconsistent data submitted by the initial carriers for each UPIN. However, rather than choosing the first practice setting regardless of DRIP status, the first *active* practice setting was chosen in order to identify a carrier that could assist in determining whether *potential* cases were actual cases or false positives. (Because the Registry is still relatively young, it is probable that a high proportion of the first practice settings are also active.) Note that the tabulations by carrier are an indication, not necessarily of the source of the multiple UPIN, but of the carrier in the best position to develop the case (with the possible exception of the two RRB carriers).

8.2 Results

The first part of this section presents the results of varying the detection algorithm. The second part presents some characteristics of multiple-UPIN cases and identifies carriers that have the highest association with multiple-UPIN cases.

[2]Providers that only had de-activated practice settings were not included in the analyses.

8.2.1 Range of Potential Multiple-UPIN Cases

The results of eight variations on the detection algorithm produced a range of 2,617 to 25,308 potential multiple-UPIN cases (first column of Table 8-1). Depending on the presence of LNAMESUM and provider credential, dropping the first initial of the first name from SCREENER identified anywhere from 3,100 to 20,000 additional potential cases. Similarly, omitting LNAMESUM resulted in 1,300 to 17,100 additional potential cases. Dropping provider credential resulted in an additional 900 to 9,000 potential cases, depending on whether LNAMESUM or the first initial of the first name was present. Inclusion of LNAMESUM in SCREENER, in general, results in losing too many good multiple-UPIN candidates.

The large range of potential cases produced by varying the detection algorithms implies that additional criteria should be utilized to assist in determining which of the algorithms is best at detecting multiple-UPIN cases. This was done by visually inspecting data elements of the first 50 potential cases produced by each of the eight algorithms. One criterion was to determine how many cases were *seemingly* false positives. Another criterion was to determine how spelling variations affected the number of potential cases.

The estimated rate of false positives ranged from 14 to 54 percent (second column of Table 8-1).[3] As expected, the rate of false positives was highest when SCREENER had the fewest components. When SCREENER relied primarily on date of birth and the first three letters of the last name, the false positive rate was up to twice as high as the other variants of the detection algorithm. The number of potential cases shown in column 1 of Table 8-1 was

[3]During our actual work, we inadvertently used the last instead of the first active practice setting for each UPIN in the eight data dumps used to estimate *seemingly* false positive rates. We compared one of them to a data dump using first active practice settings. In most cases, the first and last active practice settings were either identical or yielded the same false positive assessment. We, therefore, believe that the impact of the inadvertent use of the last active practice setting on the estimated *seemingly* false positive rate is small, especially in comparison to other factors that influence the *seemingly* false positive rate. For instance, because of the subjective nature in making false positive assessments, reasonable differences of opinion between two individuals could result in large differences in estimates of *seemingly* false positive rates.

TABLE 8-1

NUMBER OF POSSIBLE CASES WITH MULTIPLE UPINS

Screener Elements	C.A.I. Potential Number of Cases	Estimated Percent of false Positives	Net Potential Number of Cases
DOB,L1_3,F1,LNAMESUM,CREDENTIAL	2,617	16 %	2,198
DOB,L1_3,LNAMESUM,CREDENTIAL	5,702	22	4,448
DOB,L1_3,F1,CREDENTIAL	3,960	16	3,326
DOB,L1_3,CREDENTIAL	16,214	42	9,404
DOB,L1_3,F1,LNAMESUM	3,530	14	3,036
DOB,L1_3,LNAMESUM	8,201	26	6,069
DOB,L1_3,F1	5,540	20	4,432
DOB,L1_3	25,308	54	11,642

Abbreviations:

C.A.I. - Computer Algorithm Identified
DOB - Date of Birth
L1_3 - First 3 letters of last name
F1 - First letter of first name
LNAMESUM - Sum of the numbers assigned to the consonants in the last name
CREDENTIAL - Provider credential (MD, DO, DDM, DDS, DPM, OD or CH)

SOURCE: Physician Registry, July 1993.

reduced by the estimated number of false positives, to produce the final column of Table 8-1, which shows the range of the *net* potential number of cases: 2,198 to 11,642 cases. Determining whether these cases represent true multiple-UPIN providers would require carriers to: (1) check their other computerized practice setting information, (2) check internal paper (hard copy) files, and/or (3) directly contact the provider(s).

An examination of spelling variation was performed in conjunction with the identification of seemingly false positive cases. In the first instance, the spellings of the provider's last name were checked for minor discrepancies. For pairs of common last names (e.g., Anderson and Andersen), unless other information pointed to a multiple-UPIN case, the case was marked as a false positive. For very unusual last names, the absence or presence of an extra consonant (as in many Slavic names), the case would not be marked as a *seemingly* false positive. This examination helped produce an alternate assessment of the usefulness of the three optional SCREENER components. The two most important conclusions we drew from the examination were, once the computerized detection algorithms produced potential cases: (1) the results need to be visually inspected and (2) it is difficult, given the nature of the data, to incorporate other computerized data checks to reduce the number of false positives.

The above conclusions are illustrated in Exhibit 8-1. Of the five components of the SCREENER variant utilized for Exhibit 8-1, only the first initial of the first name was omitted. The UPINs of Socrates XYZ and Plato XYZ[4], who had the same DOB, credentials, and LNAMESUM, were identified by the detection algorithm as a potential case. At first glance, it would seem that the omission of the first initial produced a false positive. However, their optometry school code and graduation year were identical. Further, both the billing zip code (BILLZIP) and business zip code (BIZZIP) were very similar. Because of the possibility that they might be twins, the foregoing evidence on education and zip codes might be irrelevant. However, the state license (LICENSE) is very similar with the character string *2657P* present in

[4]Both the first and last names were changed to preserve the anonymity of the provider(s).

EXHIBIT 8-1

EXAMPLE OF A POTENTIAL MULTIPLE-UPIN CASE
(PARTIAL LISTING OF AVAILABLE DATA ELEMENTS)

NAME	DEFINITION	SCREENER ELEMENTS	UPIN-1	UPIN-2
UPIN	Unique Physician Identification Number		U X X X 0 1	U X X X 0 2
LNAME	Physicians last name	L1_B	X Y Z	X Y Z
LNAMESUM	Sum of consonants in last name	LNAMESUM	X X	X X
FNAME	Physicians first name	*F_1	S O C R A T E S	P L A T O
MNAME	Physicians middle name		E	S
SUFFIX	Physicians name suffix			
DOB	Date of birth (YYMMDD)	DOB	5 4 1 1 1 5	5 4 1 1 1 5
DOD	Date of death (YYMMDD)		0	0
MEDSCHL	The medical school code		4 0 0 0 0	4 0 0 0 0
GRADYR	Year phys. graduated from medical school		7 9	7 9
LICENSE	The state license number		O E 2 6 5 7 P 0 0 0 0 0	0 0 0 0 0 0 0 0 2 6 5 7 P
PHYSSTAT	The physician status		2	2
CREDENTL	The physicians credentials	CREDENTIAL	O D	O D
SPEC1	Physicians primary specialty		4 1	4 1
BILLZIP	The billing zip code		9 3 0 0 1 0 0 0 0	9 3 0 0 3 0 0 0 0
BIZZIP	The Business zip code		9 3 0 0 1 0 0 0 0	9 3 0 0 3 0 0 0 0
DRIP	The resident or intern indicator		P	P
SETDATE	Date the setting was added		1 9 9 1 0 1 1 7	1 9 9 1 0 2 2 6
SETLNAME	Physician last name for this setting		X Y Z	X Y Z
SETFNAME	Physician first name for this setting		S O C R A T E S	P L A T O
SETMNAME	Physician middle name for this setting		E	S
SETSUFF	Physician suffix name for this setting			
SETDOB	Physician date of birth for this setting		5 4 1 1 1 5	5 4 1 1 1 5
SETDOD	Physician date of death for this setting		0	0
SETMDSCH	Medical school code for this setting		4 0 0 0 0	4 0 0 0 0
SETGRDYR	Year phys graduated for this setting		7 9	7 9

Notes: Selected data elements changed to preserve provider anonymity.
L1_3 = First three letters of last name; LNAMESUM = sum of the numbers assigned to the consonants in the last name; *F_1 = first letter of the first name, was omitted from SCREENER in this example.

SOURCE: Physician Registry, July 1993.

both. Therefore, despite dissimilar first names, the state license indicated that this case is not necessarily a false positive, but rather, merits further investigation by the carrier.

The formats of the state license numbers for the two UPINs also illustrates the difficulties in attempting to use computerized checks of state license numbers to help detect potential multiple-UPIN cases. In one case the license number is right justified while the other is left justified (and, also, with a prefix that may or may not be an actual part of the license number).[5] (State license number also might not be of assistance when a provider is licensed and practices in different states.)

Visual examination of the potential cases revealed other difficulties that suggest that more than one algorithm should be used to detect multiple-UPIN cases. The following examples illustrate some of these difficulties.

- When an initial is used instead of a full first name, it is sometimes switched with the middle initial;
- there were instances in which the first and middle initials were present within the last name field (after the last name); and
- omitted and extra letters in the last name.

8.2.2 Selected Characteristics of Multiple-UPIN Cases

Selected characteristics of multiple-UPIN cases are presented in Table 8-2. Most of the potential cases involved multiple carriers with a range of 71 to 91 percent. (Note, however, that since only the first active practice setting was used for a given UPIN, the term *solo* does not imply that there were absolutely no other practice settings submitted by other carriers.) While most of the cases of potential multiple-UPIN providers involved physicians and osteopaths,

[5] Inconsistent license number formats seem to originate from the carriers.

TABLE 8-2

CHARACTERISTICS OF POSSIBLE MULTIPLE UPIN CASES

	SCREENER Elements				*SCREENER Elements*			
	DOB,L1_3, F1,LNAMESUM, Credential	DOB,L1_3, LNAMESUM, Credential	DOB,L1_3,F1, Credential	DOB,L1_3, Credential	DOB,L1_3, F1,LNAMESUM	DOB,L1_3, LNAMESUM	DOB,L1_3,F1	DOB,L1_3
Total Number of Multiple UPIN Cases	2,617	5,702	3,960	16,214	3,530	8,201	5,540	25,308
Multiple UPIN Cases by Presence of Multiple Carriers								
Solo Carrier	755	1,332	942	1,875	845	1,503	1,067	2,192
Multiple Carriers	1,862	4,370	3,018	14,339	2,685	6,698	4,473	23,116
Physician Status (over all UPINs)								
MD, DO	3,363	8,540	5,259	28,897	4,906	11,807	7,538	39,386
All Others	1,887	2,991	2,705	4,387	2,190	4,871	3,632	13,332
Number of UPINs per Multiple-UPIN Case								
2 UPINs	2,601	5,580	3,916	15,415	3,494	7,947	5,451	23,446
3 UPINs	16	117	44	743	36	234	88	1,650
4 UPINs	0	5	0	55	0	18	1	185
5 UPINs	0	0	0	1	0	2	0	26
6 UPINs	0	0	0	0	0	0	0	1
Number of Multiple Carrier Cases in which RRB Carriers Present								
10072 - RRB Georgia	633	831	824	1,296	941	1,190	1,167	1,872
10074 - RRB Utah	480	786	638	1,290	768	1,134	949	1,816

Notes: See Table 8-1 for definitions of SCREENER elements.
Because of inconsistencies in the value of physician status between practice settifng (Table 6-1), the total number of UPINs involved in the multipe-UPIN cases, instead of the number of cases themselves are presented in Table 8-2 for physician status.

SOURCE: Physician Registry, July 1993.

other providers were disproportionately represented in the potential cases.[6] That is, other providers account for about 11 percent of practice settings but represented from 13 to 36 percent of the UPINs in Table 8-2. The number of UPINs involved in the potential cases range from two to six, with the vast majority having two UPINs. It seems likely that false positives would account for most cases having three UPINs and all of the cases having four or more UPINs.

As has been discussed in several previous chapters, the two Railroad Retirement Board (RRB) carriers often encounter significant difficulties in obtaining information from providers. Thus, it seemed likely that the RRB carriers would be present in many multiple-UPIN cases. As can be seen in the last two rows of Table 8-2, this expectation was borne out when only the multi-*carrier* cases were considered.

To help TOLIC to detect potential multiple-UPIN cases and then to determine whether the cases are false positives or need further investigation requires the assistance of the carriers. To identify the carriers that could help investigate potential cases, the number of cases for each carrier was tabulated. For multiple carrier cases, Table 8-3 (based one of the eight SCREENER variants) shows the number of potential cases that each carrier might have to help investigate. As might be expected, the carriers with the most practice settings would tend to have the greatest investigative burden: the two RRB carriers, Michigan BC/BS, California BS, Texas BC/BS, Illinois BC/BS, and Empire BC/BS. However, the share of the cases that these carriers might have to investigate is roughly proportional to their share of Registry practice settings, except for Michigan which is over-represented. Appendix Tables E-1 through E-8 show the potential carrier burden for all eight SCREENER variants. The identity of carriers with the highest number of potential cases varies very little by the variant of SCREENER utilized.

For the so-called *solo* carrier cases, the carriers with the greatest potential investigative burden are indicated in Table 8-4 (using the same SCREENER variant as was used in Table 8-3).

[6]Because of inconsistencies in the value of physician status between practice setting (Table 6-1), the total number of UPINs involved in the multiple-UPIN cases, instead of the number of cases themselves, is presented in Table 8-2 for physician status.

TABLE 8-3

OCCURRENCE OF CARRIERS IN MULTIPLE-UPIN CASES
(SCREENER = DOB, L1_3, F1, CREDENTIAL)

Carrier Number	Carrier Name	Frequency	Percent
10072	R.R.B. - Travelers - GA	864	10.80 %
10074	R.R.B. - Travelers - UT	662	8.30
00710	Michigan BC/BS	480	6.00
00542	California BS	414	5.20
00900	Texas BC/BS	386	4.80
00621	Illinois BC/BS	335	4.20
00803	Empire BC/BS, NY	308	3.90
00865	Pennsylvania BS, PA	296	3.70
02050	T.O.L.I.C.	292	3.70
16360	Nationwide Insurance Co., OH	263	3.30
00590	Florida BC/BS	240	3.00
00860	Pennsylvania BS, NJ	214	2.70
00700	C&S Administrative Services, MA	171	2.10
00973	S.S.S., PR	145	1.80
05535	Connecticut General Life, NC	143	1.80
00630	Indiana BC/BS	142	1.80
01040	Aetna Life and Casualty, GA	134	1.70
11260	Missouri General Am. Life, MO	131	1.60
00932	Washington Physician Service	129	1.60
01370	Aetna Life and Casualty, OK	127	1.60
00801	Blue Shield of Western New York	126	1.60
10490	Travelers Insurance Co., VA	125	1.60
00528	Arkansas BC/BS, LA	112	1.40
00951	Wisconsin Physician Service	110	1.40
05440	Connecticut General Life, TN	110	1.40
00640	Iowa BS	104	1.30
10230	Travelers Insurance Co., CT	94	1.20
01360	Aetna Life and Casualty, OR	86	1.10
01030	Aetna Life and Casualty, AZ	84	1.10
00690	Maryland Blue Shield	80	1.00
00550	Colorado BC/BS	76	1.00
10240	Travelers Insurance Co., MN	73	0.90
00510	Alabama BC/BS	67	0.80
00660	Kentucky BC/BS	66	0.80
00880	South Carolina BC/BS	61	0.80
00580	Pennsylvania BS, DC	59	0.70
00720	Minnesota BC/BS	51	0.60
00740	Kansas BS, MO	50	0.60
14330	Group Health Insurance, Inc., NY	49	0.60
16510	Nationwide Insurance, WV	49	0.60
01360	Aetna Life and Casualty, NM	45	0.60
00520	Arkansas BC/BS, AR	43	0.50
00910	Utah BC/BS	36	0.50
10250	Travelers Insurance Co., MS	35	0.40
00650	Kansas BS, KS	32	0.40
00655	Kansas BS, NE	32	0.40
21200	C&S Administrative Services, ME	32	0.40
00820	North Dakota BC/BS, ND-SD	27	0.30
00870	Rhode Island Blue Shield	26	0.30
01290	Aetna Life and Casualty, NV	24	0.30
00751	Montana BC/BS	21	0.30
01120	Aetna Life and Casualty, HI	21	0.30
01020	Aetna Life and Casualty, AK	19	0.20
00570	Pennsylvania BS, DE	16	0.20
00780	C&S Administrative Services, NH	15	0.20
05130	Connecticut General Life, ID	15	0.20
00781	C&S Administrative Services, VT	13	0.20
00825	North Dakota BC/BS, WY	4	0.10

Abbreviations for SCREENER:

DOB - Date of Birth
L1_3 - First 3 letters of last name
F1 - First letter of first name
CREDENTIAL - Provider credential (MD, DO, DDM, DDS, DPM, OD or CH)

upin\tables\TAB8-3.XLS\sjb

SOURCE: Physician Registry, July 1993.

TABLE 8-4

FREQUENCY OF MULTIPLE-UPIN CASES WITH JUST ONE CARRIER, BY CARRIER
(SCREENER = DOB, L1_3, F1, CREDENTIAL)

Carrier Number	Carrier Name	Frequency	Percent
00710	Michigan BC/BS	100	10.60 %
00621	Illinois BC/BS	78	8.30
00973	S.S.S., PR	56	5.90
00900	Texas BC/BS	46	4.90
16360	Nationwide Insurance Co., OH	43	4.60
02050	T.O.L.I.C.	37	3.90
00542	California BS	34	3.60
00803	Empire BC/BS, NY	34	3.60
01370	Aetna Life and Casualty, OK	33	3.50
00590	Florida BC/BS	31	3.30
00700	C&S Administrative Services, MA	28	3.00
01040	Aetna Life and Casualty, GA	25	2.70
11260	Missouri General Am. Life, MO	25	2.70
05535	Connecticut General Life, NC	24	2.50
10490	Travelers Insurance Co., VA	23	2.40
00630	Indiana BC/BS	22	2.30
00860	Pennsylvania BS, NJ	22	2.30
00801	Blue Shield of Western New York	20	2.10
00865	Pennsylvania BS, PA	19	2.00
00932	Washington Physician Service	18	1.90
01030	Aetna Life and Casualty, AZ	18	1.90
10072	R.R.B. - Travelers - GA	17	1.80
00528	Arkansas BC/BS, LA	15	1.60
00660	Kentucky BC/BS	14	1.50
01380	Aetna Life and Casualty, OR	14	1.50
05440	Connecticut General Life, TN	14	1.50
10230	Travelers Insurance Co., CT	14	1.50
00640	Iowa BS	12	1.30
10074	R.R.B. - Travelers - UT	11	1.20
00550	Colorado BC/BS	9	1.00
00951	Wisconsin Physician Service	9	1.00
00720	Minnesota BC/BS	7	0.70
00880	South Carolina BC/BS	7	0.70
10240	Travelers Insurance Co., MN	7	0.70
00690	Maryland Blue Shield	6	0.60
01360	Aetna Life and Casualty, NM	6	0.60
00870	Rhode Island Blue Shield	5	0.50
10250	Travelers Insurance Co., MS	5	0.50
00520	Arkansas BC/BS, AR	3	0.30
00650	Kansas BS, KS	3	0.30
00655	Kansas BS, NE	3	0.30
00740	Kansas BS, MO	3	0.30
01020	Aetna Life and Casualty, AK	3	0.30
16510	Nationwide Insurance, WV	3	0.30
00510	Alabama BC/BS	2	0.20
00570	Pennsylvania BS, DE	2	0.20
00781	C&S Administrative Services, VT	2	0.20
00910	Utah BC/BS	2	0.20
14330	Group Health Insurance, Inc., NY	2	0.20
00580	Pennsylvania BS, DC	1	0.10
00751	Montana BC/BS	1	0.10
00820	North Dakota BC/BS, ND-SD	1	0.10
01120	Aetna Life and Casualty, HI	1	0.10
01290	Aetna Life and Casualty, NV	1	0.10
05130	Connecticut General Life, ID	1	0.10

Abbreviations for SCREENER:

DOB - Date of Birth
L1_3 - First 3 letters of last name
F1 - First letter of first name
CREDENTIAL - Provider credential (MD, DO, DDM, DDS, DPM, OD & CH)

SOURCE: Physician Registry, July 1993.

upin\tables\TAB8-4.XLS\wb

The five highest are Michigan BC/BS with 100 cases, Illinois BC/BS with 78 cases, S.S.S. - Puerto Rico with 56 cases, Texas BC/BS with 46 cases, and Nationwide Insurance (Ohio) with 43 cases. Again, except for Puerto Rico, these carriers are among the largest. Appendix Tables E-9 through E-16 show the potential carrier burden for all eight SCREENER variants. As with the multiple-carrier cases, identity of carriers with the highest number of cases varies very little by the variant of SCREENER utilized. Note that the RRB carriers have a disproportionately *low* share of the solo cases.

8.3 Summary

The number of potential multiple-UPIN cases ranges from 2,617 to 25,308 cases even after current detection procedures have been applied. And after accounting for seemingly false positives, the estimated number of cases ranges from 2,198 to 11,642. Of the SCREENER components, the presence of LNAMESUM results in failing to detect too many potential multiple-UPIN providers. The reason is that the omission of a consonant or presence of an extra consonant in the last name results in unequal values for SCREENER and thus a failure to detect the potential cases. Without LNAMESUM, the range of potential cases that needs to be at least visually reviewed remains large, at 3,960 to 25,308 cases. Even after accounting for *seemingly* false positives, there might be over 11,000 multiple-UPIN providers. The number of cases could be even higher if accurate dates of birth were obtained for the nearly ten percent of UPINs that have missing and/or erroneous values (Table 4-1).

The investigative burden on carriers seems to be proportionate to the number of practice settings each submitted to the Registry. Since the vast majority of cases have multiple carriers, it would require a considerable coordination effort among TOLIC and the carriers. Because the Registry's data quality dramatically improved as TOLIC and the carriers became more experienced, a one-time massive investigation of potential multiple-UPIN cases (after

improvements in the quality of date of birth data) might be sufficient to resolve most of the multiple-UPIN cases. Thereafter, normal maintenance checks would probably be sufficient to minimize the number of providers with multiple-UPINs. And the investigative burden could be minimized if it is possible to develop a flag so that previously resolved cases do not have to be re-investigated.

Use of the SCREENER variant that only includes date of birth and the first three letters of the last name yields the largest number of potential multiple-UPIN cases. With more than 25,000 cases requiring visual inspection, the investigation cost could be quite considerable. For a more restrictive SCREENER variant (e.g., the "full" 5-element variant) that yielded far fewer potential multiple-UPIN cases, the investigative cost would be considerably less. The trade-offs between investigative costs and increased accuracy need to be considered before implementing a clean-up of multiple-UPIN cases.

9.0 SUMMARY OF FINDINGS AND DISCUSSION

In response to a Congressional mandate (P.L. 99-272), the Physician Registry was established in 1989 under a HCFA contract to Transamerica Occidental Life Insurance Company (TOLIC). The primary purpose of the Registry is to issue a unique identification number (UPIN) to each provider of Medicare services. This one UPIN should then be used by the provider during his/her entire lifetime association with the Medicare program.

TOLIC processed 1.2 million provider records from 57 carriers during the first half of 1989 (the *load* phase), creating about 970,000 practice setting records for over 600,000 individual providers. It is inevitable that a task of such magnitude, performed during a short period of time, would not have been able to resolve all inconsistencies among provider records. Nor would it have been possible to correct all erroneous values or to obtain values that were missing for a data element.

Since the Registry is an on-going operation, the load phase represents but one aspect of the entire operation. From January 1, 1990, through June 20, 1993, over a million more practice settings were added to the Registry, of which nearly a third were from the two Railroad Retirement Board (RRB) carriers. The annual volume of new practice setting records and updates to existing ones greatly exceeded expectations. Data element integrity improved as Registry data edits were modified in response to the experience gained by TOLIC and the carriers in performing their Registry duties.

The major goal of this study was to document the extent to which data elements do not have valid values; to determine, if possible, the reasons why invalid values are present; and to recommend corrective methods or, where correction is not cost effective, to develop recode algorithms. In addition to checking the integrity of each data element, the consistency of data elements for a given UPIN was examined: (1) across practice settings, (2) between practice settings and the header portion of the provider's record, and (3) between different data

elements within a record (e.g., provider credential and specialty type). Lastly, estimates of the number of providers with multiple UPINs were derived. The analysis was performed on the July 1, 1993, Registry file provided by TOLIC to HCFA.

9.1 Basic Findings

The basic findings of Chapters 4 through 8 are presented in this section. The findings of Chapter 2 (from the carrier site visits) are integrated into Section 9.2.

9.1.1 Data Element Integrity

The combined rate of missing, erroneous, dubious, and unknown values (MEDU rate) was highest for both the header and practice setting versions of birth year, medical school, and graduation year - a total of six data elements. The MEDU rate was also relatively high for state license number. Missing values were the most important factor in these high MEDU rates. For the seven data elements listed above, missing value rates ranged from nine to 22 percent.

Four carriers had combined MEDU rates greater than 36 percent for the practice setting version of birth year while another 20 had combined MEDU rates between 18 and 36 percent (Figure 9-1). Five carriers had combined MEDU rates greater than 36 percent for the practice setting version of school code while another 22 had combined MEDU rates between 18 and 36 percent (Figure 9-2). Four carriers had combined MEDU rates greater than 44 percent for the practice setting version of graduation year while another 23 had combined MEDU rates between 22 and 44 percent (Figure 9-3). Seven carriers had missing value rates greater than 44 percent for state license number while another 12 had combined MEDU rates between 22 and 44 percent (Figure 9-4). For all four data elements (presented in Figures 9-1 through 9-4), the carrier for Georgia was in the group with the highest combined MEDU rates. In addition, for

FIGURE 9-1

COMBINED MISSING/ERRONEOUS/DUBIOUS RATES FOR BIRTH YEAR BY CARRIER

00801	BS of Western NY	NY					
00865	Pennsylvania BS	PA					
00590	Florida BC/BS	FL					
00780	C&S Admin. Services	NH					
10490	Travelers Ins. Co.	VA					
00528	Arkansas BC/BS	LA					
10250	Travelers Ins. Co.	MS					
00751	Montana BC/BS	MT					
00820	North Dakota BC/BS	SD-ND					
16510	Nationwide Ins. Co.	WV					
00860	Pennsylvania BS	NJ					
16360	Nationwide Ins. Co.	OH					
02050	Transamerica Occidental	So. CA					
05535	Connecticut General	NC					
00655	Kansas BS	NE					
00650	Kansas BS	KS	00630	Indiana BC/BS	IN		
00510	Alabama BC/BS	AL	00720	Minnesota BC/BS	MN		
01290	Aetna Life & Casualty	NV	00781	C&S Admin. Services	VT		
10240	Travelers Ins. Co.	MN	21200	C&S Admin. Services	ME		
01380	Aetna Life & Casualty	OR	00803	Empire BC/BS	NY		
00550	Colorado BC/BS	CO	01360	Aetna Life & Casualty	NM		
01020	Aetna Life & Casualty	AK	10074	Travelers-RRB-UT	RRB		
00951	Wisconsin Physician Serv.	WI	10230	Travelers Ins. Co.	CT		
11260	Missouri Gen. Amer. Life	MO	00910	Utah BC/BS	UT		
00870	Rhode Island BS	RI	00621	Illinois BC/BS	IL		
00710	Michigan BC/BS	MI	01370	Aetna Life & Casualty	OK		
00660	Kentucky BC/BS	KY	01030	Aetna Life & Casualty	AZ		
00710	Kansas BS	MO-KS	10072	Travelers-RRB-GA	RRB		
00640	Iowa BS	IA	00570	Pennsylvania BS	DE		
05440	Connecticut General	TN	00580	Pennsylvania BS	DC		
00973	S.S.S.	PR	00690	Maryland BS	MD		
00932	Washington Physcian Serv.	WA	05130	Connecticut General	ID	01040	Aetna Life & Casualty GA
00825	North Dakota BC/BS	WY	01120	Aetna Life & Casualty	HI	00700	C&S Admin. Services MA
00542	California BS	No. CA	00900	Texas BC/BS	TX	00520	Arkansas BC/BS AR
00974	S.S.S.	VI	14330	Group Health Ins., Inc.	NY	00880	South Carolina BC/BS SC

Less than 18 percent | 18 to 36 percent | Greater than 36 percent

Notes: This figure is based on the practice setting of birth year. For each column, the carriers with the highest values are at the top.

SOURCE: Physician Registry, July 1993.

FIGURE 9-2

COMBINED MISSING/ERRONEOUS RATES FOR SCHOOOL CODE BY CARRIER

00801	Blue Shield of Western New York	NY					
01020	Aetna Life and Casualty	AK					
00865	Pennsylvania BS	PA					
00660	Kentucky BC/BS	KY					
00710	Michigan BC/BS	MI					
00780	C&S Administrative Services	NH					
10250	Travelers Insurance Co.	MS					
14330	Group Health Insurance, Inc.	NY					
01120	Aetna Life and Casualty	HI					
00860	Pennsylvania BS	NJ					
10490	Travelers Insurance Co.	VA	01360	Aetna Life and Casualty	NM		
00528	Arkansas BC/BS	LA	01370	Aetna Life and Casualty	OK		
00740	Kansas BS	MO-KS	21200	C&S Administrative Services	ME		
00650	Kansas BS	KS	00720	Minnesota BC/BS	MN		
00751	Montana BC/BS	MT	00630	Indiana BC/BS	IN		
01380	Aetna Life and Casualty	OR	00781	C&S Administrative Services	VT		
00825	North Dakota BC/BS	WY	00803	Empire BC/BS	NY		
00550	Colorado BC/BS	CO	00820	North Dakota BC/BS	ND-SD		
02050	T.O.L.I.C.	So. CA	10230	Travelers Insurance Co.	CT		
10240	Travelers Insurance Co.	MN	01030	Aetna Life and Casualty	AZ		
00590	Florida BC/BS	FL	00932	Washington Physician Service	WA		
00510	Alabama BC/BS	AL	00621	Illinois BC/BS	IL		
00655	Kansas BS	NE	16360	Nationwide Insurance Co.	OH		
00870	Rhode Island Blue Shield	RI	00910	Utah BC/BS	UT		
05535	Connecticut General Life	NC	10074	R.R.B. - Travelers - UT	RRB		
05440	Connecticut General Life	TN	05130	Connecticut General Life	ID		
01290	Aetna Life and Casualty, NV	NV	10072	R.R.B. - Travelers - GA	RRB		
11260	Missouri General Am. Life	MO	00900	Texas BC/BS	TX	01040	Aetna Life and Casualty GA
00951	Wisconsin Physician Service	WI	00570	Pennsylvania BS	DE	00880	South Carolina BC/BS SC
00973	S.S.S.	PR	00690	Maryland Blue Shield	MD	00520	Arkansas BC/BS AR
00542	California BS	No. CA	00580	Pennsylvania BS	DC	00700	C&S Administrative Services MA
00974	S.S.S.	VI	16510	Nationwide Insurance	WV	00640	Iowa BS IA

Less than 18 percent | 18 to 36 percent | Greater than 36 percent

Notes: This figure is based on the practice setting of school code. For each column, the carriers with the highest values are at the top.

SOURCE: Physician Registry, July 1993.

FIGURE 9-3

COMBINED MISSING/ERRONEOUS/DUBIOUS RATES FOR GRADUATION YEAR BY CARRIER

00580 Pennsylvania BS	DC				
05130 Connecticut General Life	ID				
00780 C&S Administrative Services	NH				
00865 Pennsylvania BS	PA				
01020 Aetna Life and Casualty	AK				
00825 North Dakota BC/BS	WY				
00710 Michigan BC/BS	MI				
14330 Group Health Insurance, Inc.	NY				
10250 Travelers Insurance Co.	MS				
10490 Travelers Insurance Co.	VA	00720 Minnesota BC/BS	MN		
00751 Montana BC/BS	MT	00520 Arkansas BC/BS	AR		
00974 S.S.S.	VI	00700 C&S Administrative Services	MA		
01120 Aetna Life and Casualty	HI	00820 North Dakota BC/BS	ND-SD		
16510 Nationwide Insurance	WV	01360 Aetna Life and Casualty	NM		
00528 Arkansas BC/BS	LA	00781 C&S Administrative Services	VT		
00650 Kansas BS	KS	01370 Aetna Life and Casualty	OK		
00860 Pennsylvania BS	NJ	21200 C&S Administrative Services	ME		
10240 Travelers Insurance Co.	MN	00910 Utah BC/BS	UT		
00590 Florida BC/BS	FL	16360 Nationwide Insurance Co.	OH		
00655 Kansas BS	NE	10230 Travelers Insurance Co.	CT		
01380 Aetna Life and Casualty	OR	00621 Illinois BC/BS	IL		
00550 Colorado BC/BS	CO	00803 Empire BC/BS	NY		
00870 Rhode Island Blue Shield	RI	10074 R.R.B. - Travelers - UT	RRB		
11260 Missouri General Am. Life	MO	00932 Washington Physician Service	WA		
00510 Alabama BC/BS	AL	01030 Aetna Life and Casualty	AZ		
01290 Aetna Life and Casualty	NV	10072 R.R.B. - Travelers - GA	RRB		
00660 Kentucky BC/BS	KY	00900 Texas BC/BS	TX		
00740 Kansas BS	MO-KS	05535 Connecticut General Life	NC		
00640 Iowa BS	IA	00570 Pennsylvania BS	DE	00951 Wisconsin Physician Service	WI
05440 Connecticut General Life	TN	00690 Maryland Blue Shield	MD	01040 Aetna Life and Casualty	GA
00542 California BS	No. CA	00630 Indiana BC/BS	IN	00880 South Carolina BC/BS	SC
00973 S.S.S.	PR	00801 Blue Shield of Western New York	NY	02050 T.O.L.I.C.	So. CA
Less than 22 percent		22 to 44 percent		Greater than 44 percent	

Notes: This figure is based on the practice setting of graduation year. For each column, the carriers with the highest values are at the top.

SOURCE: Physician Registry, July 1993

FIGURE 9-4

MISSING VALUE RATES FOR STATE LICENSE NUMBER BY CARRIER

10250	Travelers Ins. Co.	MS					
10240	Travelers Ins. Co.	MN					
00865	Pennsylvania BS	PA					
10490	Travelers Ins. Co.	VA					
01290	Aetna Life & Casualty	NV					
16510	Nationwide Ins. Co.	WV					
00801	BS of Western NY	NY					
00528	Arkansas BC/BS	LA					
00803	Empire BC/BS	NY					
00650	Kansas BS	KS					
00655	Kansas BS	NE					
00860	Pennsylvania BS	NJ					
16360	Nationwide Ins. Co.	OH					
00910	Utah BC/BS	UT					
00751	Montana BC/BS	MT					
00870	Rhode Island BS	RI					
00740	Kansas BS	MO-KS					
05440	Connecticut General	TN					
14330	Group Health Ins., Inc.	NY					
05535	Connecticut General	NC					
00621	Illinois BC/BS	IL					
00951	Wisconsin Physician Serv.	WI					
00720	Minnesota BC/BS	MN					
00590	Florida BC/BS	FL					
00820	North Dakota BC/BS	SD-ND					
00550	Colorado BC/BS	CO					
00640	Iowa BS	IA					
00880	South Carolina BC/BS	SC					
00932	Washington Physician Serv.	WA	01380	Aetna Life & Casualty	OR		
00973	S.S.S.	PR	00780	C&S Admin. Services	NH		
00510	Alabama BC/BS	AL	01370	Aetna Life & Casualty	OK		
00542	California BS	No. CA	01360	Aetna Life & Casualty	NM		
00630	Indiana BC/BS	IN	10074	Travelers-RRB-UT	RRB		
00660	Kentucky BC/BS	KY	00690	Maryland BS	MD	00781	C&S Admin. Services VT
00710	Michigan BC/BS	MI	01030	Aetna Life & Casualty	AZ	10230	Travelers Ins. Co. CT
00825	North Dakota BC/BS	WY	10072	Travelers-RRB-GA	RRB	00700	C&S Admin. Services MA
00900	Texas BC/BS	TX	01020	Aetna Life & Casualty	AK	21200	C&S Admin. Services ME
00974	S.S.S.	VI	00570	Pennsylvania BS	DE	00520	Arkansas BC/BS AR
02050	Transamerica Occidental	So. CA	01120	Aetna Life & Casualty	HI	11260	Missouri Gen. Amer. Life MO
05130	Connecticut General	ID	00580	Pennsylvania BS	DC	01040	Aetna Life & Casualty GA

Less than 22 percent | 22 to 44 percent | Greater than 44 percent

Note: For each column, the carriers with the highest values are at the top.

SOURCE: Physician Registry, July 1993.

upin\FIG9-4.XLS\sjb

three of the four data elements the carriers for Arkansas, Massachusetts, and South Carolina were in the groups with highest combined MEDU rates.

The MEDU rates for the seven data elements were highest during the year the Registry began operations (1989) and declined thereafter, with lower missing value rates accounting for most of the decline. For the practice setting versions of birth year, medical school, and graduation year, persisting difficulties confronting the Railroad Retirement Board (RRB) carriers have prevented the national missing value rates from falling below five to six percent in recent years (1992 and 1993). For instance, 98.6 percent of the 23,206 missing values for birth year in 1992-93 were attributable to the RRB carriers. For all other carriers combined, the missing value rate for birth year was 0.1 percent.

All types of providers had MEDU rates far in excess of the 0.5 percent standard (employed in the analysis) for each of the seven data elements listed above. Doctors of medicine had the highest MEDU rates for birth year (both header and practice setting versions), medical school (both header and practice setting versions), and graduation year (practice setting version). Chiropractors, podiatrists, and optometrists had the highest MEDU rates for graduation year (header version), while dentists (either DDS or DDM credentials) had the highest MEDU rates for state license number.

9.1.2 Logical Inconsistencies Between Data Elements

Most of the inconsistent values between pairs of data elements (e.g., school type and provider credential) is due to missing information (i.e., one value is present and one is missing). This is especially evident with regard to sanction dates, length of sanction, and school code. In turn, most of the missing values seem to be associated with practice settings added to the Registry during 1989. The missing value problem with regard to the sanction variables appears to be different from the other data elements. It seems that if carriers knew a

provider was sanctioned prior to the Registry's implementation, they did not always know the sanction date nor the length of sanction. (The codes currently used by the Registry that can be used for sanction date and length of sanction do not allow differentiation between *no sanction* and *sanctioned, but missing sanction date or missing length of sanction.*)

Despite the above difficulties, the coding of variables was usually consistent with the coding of other variables in the overwhelming majority of cases.

9.1.3 Consistency Across Practice Settings

More than ten percent of UPINs with multiple active practice settings had school codes that were not identical across practice settings. For graduation year and full sanction date, the findings were similar.[1] Additionally, about seven percent of UPINs had inconsistent dates of birth and over two percent of UPINs had inconsistent sanction codes (feasible, not blank-filled), lengths of sanction (feasible values grater than zero), and provider credentials. Exclusion of the RRB carriers usually reduced the rates by a third to a half. Multiple carriers for a given UPIN was not the sole reason why inconsistent values existed. For school codes, for instance, over twenty carriers had inconsistency rates exceeding five percent of feasible values for multiple active practice settings serviced within the carrier.

9.1.4 Consistency Between Common Header and Practice Setting Data Elements

In comparing values of the header version of data elements to their practice setting counterparts, only feasible values were included in the comparison. The inconsistency rates ranged from three to five percent for date of birth, school code, and graduation year. However, when restricting comparisons to feasible values greater than zero or non-blank, the

[1]Only feasible values of the data elements were compared.

inconsistency rates for the sanction variables all increased, with about five percent for sanction date being the highest.

9.1.5 Providers with Multiple UPINs

We estimated that the number of potential multiple-UPIN cases ranges from 2,198 to 11,642 cases. (Before accounting for seemingly false positives, the estimated number of cases ranges from 2,617 to 25,308.) Many of the potential cases are due to slight spelling variations of provider last names or extraneous characters following the last name. That is, in the presence of an otherwise identical or near-identical set of data elements on another UPIN record, a minor spelling variation is reason to suspect that a provider was assigned more than one UPIN. Because medical doctors and osteopaths comprise the bulk of the Registry data base, most of the cases of potential multiple-UPIN providers involved doctors of medicine and osteopathy; however, other types of providers were disproportionately represented in the potential cases.

A labor-intensive review would be required to determine the actual number of multiple-UPIN providers. Even after accounting for all of the false positives, the number of multiple-UPIN providers could very well exceed 11,000. The number of cases could be even higher if accurate dates of birth were obtained for the nearly ten percent of UPINs that have missing and/or erroneous values. If the actual number of providers with multiple UPINs is eleven thousand, then about 1.5 percent of providers have multiple UPINs.

9.2 Discussion

The highest MEDU rates were for birth year, school code, graduation year, and state license number. To a large extent, these relatively high rates can be readily explained by

restrictions placed on the carriers during the implementation phase (1989) of the Registry. For instance, carriers were not allowed to directly contact providers to obtain information that was not present in their provider files. Further, school code, graduation year, and state license number were not required data elements before 1992. The restrictions and changes in data element requirements account for the relatively high missing value rates, especially for 1989-91.

The RRB carriers continue to experience difficulties in obtaining values for date of birth, school code, graduation year, and state license number. Unlike typical carriers, the RRB carriers service provider claims from almost every state. Additionally, RRB beneficiaries often represent a small proportion of providers' Medicare case load. The result is that providers often do not respond to RRB requests for information, even when informed that non-response would result in the claim not being paid. The RRB carriers, in turn, are not allowed to obtain the requisite information from the Registry. Given the difficulties the RRB carriers have in obtaining accurate provider information, their role in Registry operations might warrant reconsideration.

Of the seven data elements that have high MEDU rates, the most important is date of birth. Date of birth is critical to the process of identifying providers and insuring that providers are issued only one UPIN. Another data element that might assist carriers and TOLIC identify providers is state license number. In addition to high missing value rates for state license number, the internal format of the data element is often not consistent within carrier (see Chapter 8).

To clean up date of birth alone could involve contacting as many as 64,000 providers -- almost 10 percent of providers on the Registry. Obtaining missing state license numbers and fixing the internal format might be an even more formidable task.

It is not clear to what extent carriers would be able to clean up these two data elements, let alone any others, using provider information that they currently possess but that is not fully integrated into their automated provider files. Additionally, carriers often have difficulty in

obtaining information from providers once PINs have been issued because the carriers do not have the leverage to secure provider cooperation (e.g., carriers find it more difficult to threaten providers with non-payment of claims).

The benefits to cleaning date of birth and the other data elements fall into two primary categories: (1) operations and (2) research. For operations, the primary benefit of more accurate information would be the reduction in the number of providers with multiple UPINs. Fewer multiple-UPIN providers makes it easier to conduct administrative review of providers that might not be honoring their agreed-to Medicare assignment status. Administrative review of physician compliance with HCFA ownership rules regarding investments in laboratories would be easier to conduct if there were fewer multiple UPIN cases. Reduction in the number of providers with multiple UPINs also makes it easier for PROs and other administrative users of the Registry to identify specific providers (e.g., to identify sanctioned providers). Moreover, the intent of the enabling legislation was to uniquely identify physicians; thus an individual should have only one number. More accurate dates of birth also would facilitate review of elderly providers who should perhaps have the surrogate UPIN for retired providers rather the standard UPIN. As time goes on, there will be more provider relocations and marriages/name changes in the history of the Registry file, so that the problem of multiple UPINs, to the extent it is due to missing dates of birth, could get even worse. In this light, cleaning up the key discriminator variables (e.g., date of birth and state license number) should be done as early as possible.

For research, fewer multiple-UPIN providers increases the accuracy of studies that aggregate utilization and charges to the individual provider level. Fewer multiple-UPIN providers also increases confidence in studies that link the provision of ancillary services to the referring (ordering) physician. Increased quality of variables such as graduation date, that have high missing value rates, allows better studies of the effect of provider experience on the provision of services to Medicare beneficiaries. In general, higher quality data assists research

because fewer provider records need to reviewed to determine whether they should be discarded. Fewer discarded records, in turn, implies that there are fewer biases (e.g., geographic) in provider samples drawn from the Registry file.

REFERENCES

Health Care Financing Administration, Medicare Carriers Manual- Professional Relations, Department of Health and Human Services, Part 4, Transmittals 1-7, April 1989- May 1993.

Health Care Financing Administration, Program Memorandum to Carriers, Department of Health and Human Services, March 1992.

Health Care Financing Administration, Program Memorandum to Carriers 91-B-12, Department of Health and Human Services, November 1991.