



Health Economics Resource Center

Technical Report # 8

August 26, 2003

## **A Comparison of the National VA Outpatient Database to Electronic Medical Records**

Lenny Ayyangar, MS

Jodie Trafton, PhD

Paul Barnett, PhD

### **Overview**

This technical report compares two sources of information on ambulatory care provided by the U.S. Department of Veterans Affairs (VA). We compared the VA electronic medical record to the VA national outpatient utilization database. We wished to validate data to be used in the economic component of the Multisite Opioid Substitution Treatment (MOST) study. The MOST study is evaluating the effect of adherence to clinical practice guidelines on the cost and outcomes of patients being treated for opiate dependence at seven sites.

We looked at a random sample of ambulatory care data for a small sample of patients. We compared the VA medical record, VISTA,<sup>1</sup> to the outpatient events file, a SAS extract of the National Patient Care Database (NPCD).

### **Methods**

We evaluated data on outpatient visits that occurred in the fiscal year ending on September 30, 2001. We took a random sample of three subjects enrolled at each of six MOST study sites, a total of 18 patients. We randomly sampled dates on which these subjects received outpatient care in each database. For each subject, we selected 10 dates from the VISTA database and 10 dates from the NPCD database. We then gathered data on outpatient encounters that occurred on the dates of these 20 visits from both the VISTA and the NPCD databases.

We extracted VISTA data using the outpatient encounter ad hoc report within the Health Summary package. Data displayed on the computer terminal were pasted into a text file and then read into a SAS dataset. We did not extract the VISTA laboratory or radiology packages.

We used the outpatient events SAS extract from the NPCD, an electronic file located at the national VA computer center in Austin, Texas. We selected all records for the patient sample for

---

<sup>1</sup> VISTA, the Veterans Integrated Health Systems Technology and Architecture. Its clinical applications are known as the CPRS, Computer Patient Record System.

the selected dates. To make the NPCD data consistent with the VISTA data, we excluded visits to clinic stop 108 (laboratory) and clinic stop 105 (radiology).

### Comparison of VISTA and NPCD Visit Dates

We randomly sampled 10 dates for 18 patients from each of the two sources. There were 180 combinations of date and patient selected from VISTA, and 180 selected from NPCD. Of these combinations, 27 were selected from both data sources. Thus, there were 333 unique combinations of date and patient, 153 selected from VISTA, 153 selected from NPCD, and 27 selected from both VISTA and NPCD (Table 1).

**Table 1**  
**Sample of Visit Dates from VISTA and NPCD**

Number of Dates	Source Data
153	Dates sampled in VISTA
153	Dates sampled in NPCD
27	Dates sample in both VISTA and NPCD
333	Total number of dates sampled

For one of the dates selected from NPCD, there was no corresponding visit in the VISTA (Table 2). For two of the dates selected from VISTA, there was no corresponding visit in the NPCD file.

**Table 2**  
**Correspondence Between Dates of Care Recorded in VISTA and NPCD**

Number of Dates	Correspondence
2	Dates with care found only in VISTA
1	Dates with care found only in NPCD
330	Dates with care found in both VISTA and NPCD
333	Total number of dates sampled

### Comparison of VISTA and NPCD CPT Codes

We compared the Current Procedure and Terminology (CPT) codes used to characterize care in each of the data sets. We reorganized the VISTA and NPCD data so that there was one database record for each unique combination of patient, visit date, and CPT code.

The creators of NPCD have adopted a programming rule that a given CPT code can be used only once to characterize the care given to a specific patient in a given clinic on a particular date. When the NPCD is extracted to create the outpatient events file, duplicate uses of a CPT code in

a single visit are eliminated. We therefore summarized the VISTA files in a similar way, to be consistent with the NPCD outpatient SAS extract. We dropped 21 duplicate procedure codes from the VISTA data, or 3.6% of the total.

There were 581 combinations of patient, date, and CPT code in the two files. Of these, 574 (99%) appeared in both VISTA and NPCD files (Table 3).

**Table 3**  
**Correspondence Between Procedures Recorded in VISTA and NPCD**

Number of Procedures	Correspondence
6	Procedures found only in VISTA
1	Procedures found only in NPCD
574	Procedures found in both VISTA and NPCD
581	Total procedures

There was one combination (0.2%) found only in NPCD. This was the date sampled in NPCD that had no visit recorded in VISTA. The CPT code in the NPCD record J8499, dispensing of an oral prescription (non-chemotherapy). We examined the VISTA database for this patient. It indicated that the patient failed show up for an appointment.

There were six procedures (1%) found only in VISTA (Table 4). The procedures were provided in four visits. Two of these visits (numbers 1 and 2) had a corresponding record in NPCD, but the VISTA record had an additional CPT code. The remaining two visits (numbers 3 and 4) had no corresponding record for the care in NPCD.

**Table 4**  
**Procedures Recorded in VISTA But Not in NPCD**

Visit	Procedure	Explanation
1	90853 Group psychotherapy	CPT code not found in NPCD
2	90853 Group psychotherapy	CPT code not found in NPCD
3	99212 Office/Outpatient Visit	Visit not found in NPCD
4	86485 Skin test for yeast infection	Visit not found in NPCD
4	86580 Skin test for tuberculosis	Visit not found in NPCD
4	90862 Medication management	Visit not found in NPCD

### **Comparison of VISTA and NPCD Provider Fields**

VISTA records the names of providers involved in each encounter. The NPCD uses an identification code to characterize providers. We wished to determine if the information in the two sources was consistent.

This effort was aided by the release of a national database of providers in April, 2003. The database includes the NPCD provider identification number and the first and last names of each provider. We assigned the NPCD provider identification number to all providers listed in the VISTA data. We looked up this number in the provider database using the three-digit medical center identifier number (STA3N) and the providers' first and last names. An important limitation of our analysis was the use of a 2003 provider database to study care provided in 2001. The database excluded providers who had left VA between 2001 and 2003.

Of the 91 unique provider names in the VISTA encounter data, we found 77 names in the national provider identification file. The 14 providers whose names could not be matched took part in 39 encounters. We assigned provider identification numbers to them based on an implicit match to the NPCD data. We searched in NPCD for encounters having the same patient identifier, visit date, and CPT code. We then found the provider identification number for the encounter and assigned it to the provider name. For each name we implicitly associated with an identification number, we checked that whenever that name appeared as the provider in VISTA, there was a corresponding record in NPCD with that provider's implicitly assigned identification number. We judged correspondence by an exact match of the patient identifier, visit date, and the CPT codes assigned.

Of the 14 providers in VISTA who were not found in the provider identification database, we were able to implicitly assign a provider identification to 11. This left 3 providers to whom we could not assign identification numbers. One provider was associated with a visit that appeared in VISTA but not in the NPCD. The remaining two providers were associated with a single visit found in both VISTA and NPCD. They appeared nowhere else in the data, and so it was impossible to implicitly match the names to an identification number.

We were thus able to assign 88 provider identification numbers to the providers listed in VISTA, 77 from the provider database, and 11 by implicit matching to the NPCD data. We created a view of the VISTA data that had one record for each encounter with a provider that occurred with a particular subject and a particular date.

We found 87 different provider identification numbers in the NPCD encounters in this study. We created a similar view of the NPCD data. We arranged this data so there was one record for each encounter with a provider that occurred with a particular subject and a particular date.

We compared these VISTA and NPCD data (Table 5). There were 527 encounters with providers. 512 of the encounters were recorded in both files. There were 3 provider encounters found only in NPCD. There were 12 encounters found only in VISTA. Among these 12 encounters were the previously described 5 encounters, each involving 3 providers, who could not be implicitly assigned a provider identification number.

**Table 5**  
**Comparison of Provider Encounters in VISTA and NPCD**

Number of Encounters	Explanation
512	Encounters with same provider in both VISTA and NPCD
3	Encounters with provider only in NPCD
12	Encounters with provider only in VISTA
527	Total encounters with providers

### **Definition of an NPCD Record**

We explored the definition of a record in the NPCD database. The NPCD characterizes the type of outpatient care using a three-digit code called a *clinic stop*. An NPCD record is created each time a patient visits a clinic stop on a given day. Some NPCD records have multiple providers; the database records up to 10 provider identification numbers per visit. For other visits, there are multiple NPCD records sharing the same clinic stop and visit date, each with a different provider. We were interested in learning whether VISTA data could explicate differences in the way care is characterized in the NPCD database.

There were 68 records in NPCD that shared the same patient identifier, visit date, and clinic stop. We found five records that also shared the same CPT code. When we examined these visits in VISTA, we found exactly the same information. Thus, VISTA data were unable to help diagnose the cause of duplication across the 68 NPCD records.

### **Comparison of Providers and Clinic Stops in the NPCD File**

A cross-tabulation of providers and clinic stops in the NPCD file shows that there are some providers associated with a particular clinic stop. For example, a frequency count of providers for clinic stop 523 shows 8 provider IDs appearing 10 or more times for this particular clinic stop.

### **Conclusions**

We found a high correlation between VISTA and the SAS outpatient extracts of NPCD. About 1% of encounters had different dates in the two sources, and about 1% had different procedures recorded. We found nearly as high concordance between providers names recorded in the two data sets.

Possible explanations for the differences that we observed include the following:

Methodology. We used the outpatient encounter package in VISTA. We used the SAS outpatient extract from NPCD, exclusive of clinic stops 105 and 108. These data sources may not be exactly comparable.

Differing Time Frame. We extracted VISTA data in 2003. Outpatient data were transmitted to NPCD shortly after each visit. The NPCD extracts also reflects changes transmitted to NPCD through December 2001. Subsequent changes to the data made in VISTA would not be reflected in NPCD.

VISTA Data Entry Problems Another possible source of the differences is record entry problems in VISTA . An incomplete record entry could create a record in VISTA but not in NPCD.

We found that the NPCD is a reasonably accurate representation of data in VISTA, but that some data found in VISTA were not recorded in NPCD. This included a cancelled visit that was nevertheless transmitted to NPCD, and other visits that were recorded in VISTA but did not appear in NPCD.