Adding Value to Your Data

Strategies for improving the consistency, reliability, and utility of your HMIS data.

Fred Berman, Cambridge Continuum (617-349-6209 / fberman@cambridgema.gov)

A couple of examples…

To what extent can the data challenges illustrated by the examples be resolved …

- By improving the reliability of the data?
- By strengthening the de-duplication algorithm?
- By clarifying the meaning of questions and/or the choices available for answering those questions?
- By adding a narrative context?
- By supplementing data collection to include fields that refine our understanding of the clientele or program outcomes?

Steps we can take to add value…

Basic Steps

- User Training
- Ongoing data quality review, feedback, and user support
- Modifications to data collection procedures to simplify data entry and/or reduce the incidence of blanks, "Don't Know", "Refused" … without increasing guessing or default “No”s when staff can't obtain reliable answers.
  - Pre-Homeless Zip Code: if you allowed staff to find zipcodes in an easy-to-access list of communities, would you get better answers?
  - Do you ask about Disability generically first and then drill down for specific disabling conditions, or do you ask about each specific disabling condition first, and only ask about generic disability if all the answers regarding specific disabling conditions are negative?

Steps we can take to add value…

Training to ensure that staff throughout the CoC all share the **same interpretation of questions and answer choices**; and/or **modifying the wording of the questions or options** or insertion of **explanatory text** to clarify intent of questions/choices

- Do staff know the definition of chronic homelessness?
- Do staff understand the difference between a chronic illness/health problem and a "developmental disability"?
- Are Medicaid and TANF called by different names in your community? Are there other sources of income/benefits that need to be listed so that staff don't have to remember to list them under "Other"?
- Prior Living Situation: Apartment or house that you own → Client-Owned Dwelling (condo/apartment or house)

Steps we can take to add value…

Strategies to ensure **consistent identification of clients** (e.g., ID swipe cards) and use of a **robust de-duplication** algorithm that can "overlook" typos in 1-2 identifiers if enough of the other identifiers reliably indicate a match.

- In analyzing the 2009 Cambridge CoCs AHAR data set, had we required matches of all four fields used as identifiers (First Name, Last Name, DOB, Last Four Digits of SSN), we would have only found 42% of the duplicate identities. That is, 58% of duplicate identities would have been treated as a separate client.
  - Note that forming a **composite identifier** using parts of these four basic identifiers, and then looking for matching composite identifiers would be **more or less the same as requiring all four basic identifiers to match**. Such an algorithm would have found **fewer than half of all duplicate identities** in our data set. For more info, see www.hmis.info/Resources/305/Technical-Guidelines-for-Unduplicating-and-De-identifying-HMIS-Client-Records.aspx

Steps we can take to add value…

Robust De-duplication (cont.)

In analyzing the 2009 Cambridge CoCs AHAR data set…

- If we had required matches of all properly answered identifiers (fields with meaningful data, i.e., excluding blanks, "Unknown", and "Refused" in First Name, Last Name, DOB, Last Four Digits of SSN), we would have found 62% of the duplicate identities (if we required matches in at least two fields). That is, 38% of duplicate identities would have been treated as distinct clients.
- If we had considered matches to have been found when at least two out of four identifier fields matched (i.e. over-looking mis-spelled names or miscopied SSNs or DOBs, we would have found 99% of matches (but in a handful of cases, where just the first and last names matched, we would have incorrectly assumed that the two clients were the same person).

Make sure your algorithm doesn't use matches between Blanks, "Unknown"s, or "Refused"s as a basis for identifying duplicates.

Steps we can take to add value…

Exploiting the capacity of the software to catch potential data entry errors **<u>during</u>** data entry:

- Eliminating defaults to “today's date” in programs in which staff data entry routinely occurs several days after the fact
- Range verification for dates to ensure that one-day stays are not accidentally entered as one month stays or one year stays
- Real-time feedback to the person entering data about the possibility that clients being added to the data base are already there. **<u>Err on the side of caution</u>**, i.e., use a match-finding algorithm that errs on the side of generating false positives rather than false negatives: it’s better to “annoy” the user with prompts asking whether a new client is the same as a person already in the system, than to have to remove duplicates later because staff weren't adequately warned by the system….

Steps we can take to add value…

Exploiting the capacity of the software to catch potential data entry errors **<u>during</u>** data entry *(cont.)*:

- Real-time feedback about probable inconsistencies, for example:
  - ✓ Client receives SSDI or other disability benefit, but is not identified as having a disability
  - ✓ Client fled domestic violence, but is not identified as having a recent history of domestic violence
  - ✓ Client receives veterans' benefits or SSI, but is not listed as having health coverage
  - ✓ Client came from transitional housing, but is listed as being chronically homeless
  - ✓ Client’s Housing Status is “literally homeless,” but their prior living situation is typically not considered “homeless”

Steps we can take to add value…

- Follow Barbara Ritter's advice and provide staff with (easily to access or periodically furnished) graphic summaries of the data they've collected. That is, make the data meaningful to staff.
  - If the data doesn't match what the staff know about their clients, they'll be more motivated to fix it. (Speaking from experience, don't wait til the APR or AHAR is due to do this….)
  - If supplemental data tells a more "relevant" story about the challenges of the caseload or program accomplishments, staff will be more inclined to sustain that data entry.

(Confession: we don't do nearly enough of this in the Cambridge CoC.)

- Interview program staff to find out how their different program models impact data collection (e.g., ability to observe clients in a scattered site setting vs. in a congregate setting, ability/willingness to collect data about domestic violence or HIV/AIDS, etc.)

Steps we can take to add value…

An Example: Using the Arizona Matrix

A Focus Group of family shelter staff surfaced a lack of clarity / consistency among case managers about the meaning of scoring categories in "straightforward" domains (e.g., income, child care, adult education) as well as in more "subjectively" scored domains (e.g., life skills, conflict resolution, parenting skills).

To what extent could these data challenges be resolved …

- By training?
- By clarifying the meaning of questions and/or the choices available for answering those questions?
- By adding a narrative context?
- **What other strategies would you suggest?**

Steps we can take to add value…

<u>Supplemental Data Collection to Fill Information Gaps</u>

- **Reasons for supplemental data collection**
  - Fill information gaps about clients (e.g., why they are at-risk or became homeless, obstacles faced, challenges posed to programs seeking to end their homelessness, etc.)
  - Assemble data needed to assess efficacy of program interventions, i.e., describing the interventions, describing client circumstances that might explain different outcomes with respect to targeted domains or with respect to housing / income / employment /self-sufficiency goals
  - Track performance measures of interest to funders
  - Track accomplishments too “small” to interest funders, but that signify real progress for clients, and give staff a chance to see tangible successes
  - Replace paper record-keeping
  - Build capacity to complete quantitative analyses of clientele and comparative analyses of program/agency/Continuum performance

Steps we can take to add value…

<u>Supplemental Data Collection to Fill Information Gaps</u> *(cont.)*

- **Cautions about supplemental data collection**
  - Beware of overloading staff or asking staff to collect data that they realistically aren't situated/qualified to collect
  - Make sure you know how you want to use the data you collect, and design the fields so that they are suited to that purpose:
    - If you hope to use the data to <u>supplement or replace paper files</u>, then a <u>small number</u> of narrative questions may be appropriate
    - If you hope to use the data for <u>quantitatively</u> describing the clientele, the use of program services, or client outcomes, then data fields <u>must</u> be in a format that can be aggregated: yes/no, multiple choice, scoring using a prescribed scale
  - Involve program staff in framing the questions and the choice options, so that they make sense to users.

Steps we can take to add value…

<u>Supplemental Data Collection to Fill Information Gaps</u> *(cont.)*

- **A Hypothetical Example:**
  - Tracking Employment History to Supplement Paper Files:

    Provide the following information for all employment over the past two years:
    - ✓ Name and location of employer
    - ✓ Position title and hourly/monthly salary
    - ✓ Start and end dates
    - ✓ Reason for leaving (selected from a pick list)
  - Tracking Employment History to Categorize Clients:
    - ✓ 1 = Employed more or less continuously one year or longer
    - ✓ 2 = Employed for at least six months over the past year
    - ✓ 3 = Employed for at least three months over the past year
    - ✓ 4 = Employed intermittently or for less than a total of 90 days over the past year
    - ✓ 5 = Previously employed, but not for at least a year
    - ✓ 6 = Never employed

Steps we can take to add value…

<u>Supplemental Data Collection to Fill Information Gaps</u> *(cont.)*

- **A Few More Considerations:**
  - One size doesn't fit all.
  - Standardizing vs. customizing – tailoring fields and choices to the program vs. retaining the ability to aggregate data across programs
  - Organization and Frequency of data collection – grouping fields that make sense to collect at the same time; ordering the fields so that there is a logical flow to the questions
  - Separate data collection instrument vs. Integrating questions into existing data collection instruments – same considerations about frequency of data collection and logical flow

Steps we can take to add value…

Summary

1. Work to improve the reliability of the data through
   - ✓ training,
   - ✓ staff-informed re-framing of questions and/or pick-list choices,
   - ✓ insertion of clarifying text,
   - ✓ software customizations to provide real-time feedback to users about possible data entry problems (e.g., duplicate entry of existing clients, inconsistent answers, etc.),
   - ✓ generating easily accessible reports that graphically reflect the caseloads and work done by staff.
2. Implement carefully targeted supplemental data collection in a way that minimizes additional burden to staff and blends as smoothly as possible with ongoing work.

Steps we can take to add value…

3. **Ensure that someone is regularly reviewing data for quality and completeness, providing feedback to staff about any data entry problems, and providing support to users who need it.**

-- Good Luck! --