Chapter 18
Data Management and Commingled Remains at Mass Fatality Incidents (MFIs)

Michael Hennessey

Introduction

Commingling occurs when the remains of more than one victim become intermixed. A thorough review of fragmentary remains during triage at the morgue should resolve most instances of commingling, but in some cases commingling may be detected only when separate examinations yield contradictory results. For example, at the World Trade Center (WTC), multiple DNA tests on the same set of remains sometimes yielded different profiles. In other cases, dental X-rays and DNA from the same set of remains were linked to different victims. Conflicting antemortem and postmortem data can also indicate commingling. In another case from WTC, the DNA from a victim's reference sample (known personal effect) matched a set of remains, yet the jewelry on the remains was contrary to the description provided by the spouse. After further review, the identification was reassigned. In each of these cases, visual examination had not initially detected the presence of commingled remains.

A key lesson learned from mass fatality incidents (MFIs) is that one cannot rely on a single identification method to detect commingling. All postmortem results for a case need to be reviewed for consistency. Then all of the postmortem data must be reconciled against all of the antemortem information associated with the identification. This administrative review process ("admin review") relies on being able to locate all of the case information in a timely fashion. Unfortunately, gathering all of the case information is error-prone as missing data are not always obvious. For example, if there is a matching DNA result, one might not think to ask if there is a *second* DNA test extant.

The main point of this chapter is that even if great care is exercised in the recovery and analysis of remains, the identification process may be compromised if the data are not managed correctly and an investigator misses an inconsistency. This chapter will examine the most common challenges in this regard and suggest some solutions. The framework presented here is that admin review is a data management process that spans the antemortem and postmortem information gathering systems and unites the various identification methods (anthropology, DNA,

From: *Recovery, Analysis, and Identification of Commingled Human Remains*
Edited by: B. Adams and J. Byrd © Humana Press, Totowa, NJ

fingerprint, dental, pathology, property, and X-ray). In this way, the detection and resolution of commingled remains is a multidisciplinary effort.

Unless otherwise noted, the examples in this paper are real cases. The names and identifying numbers have been anonymized to protect the privacy of the families.

MFI Data

In response to an MFI, a family assistance center (FAC) is usually established to provide services to the families of the victims. In addition, interviewers at the FAC will collect missing persons reports. Employers will provide information about staff presumed lost at the scene, and other organizations (tour agencies, airlines) may also have data to contribute regarding the victims. Anyone filing a report is an *informant,* and this antemortem (AM) information can be divided into three categories:

1. Primary data for identifications: Fingerprints, X-rays (primarily dental), and DNA are the principal methods of identification. Although other techniques can be used, such as tracing the serial number from a surgical implant, they are not applicable in most cases. While families do not always bring the X-rays or fingerprint cards to the FAC, they do provide the contact information for the custodians of this data. DNA reference materials (such as a razor or toothbrush), on the other hand, are usually provided directly by the family.
2. Secondary data for identifications: Although not sufficient to make a positive identification on its own, descriptive information about the victim's jewelry, clothing, tattoos, surgical histories, or distinctive anatomical features can be used to generate investigational leads that can point to the correct primary data for an identification. Secondary data can also be used to return property to the families of the victims.
3. Administrative data for case management: The victim's contact information (name, date of birth, address, phone number, etc.) and demographic data (age, race, gender, stature, etc.) are used to create and manage the case file. The demographic data can also contribute to the creation of investigational leads as well as the detection of potential problems (e.g., a male DNA profile associated with female remains). In addition, a case number is usually established at this time. Information about the interview itself belongs in this category: the informant and interviewer's name, their contact information, date and time of the interview, etc.

When remains are received in the morgue, they will generate postmortem (PM) data that are also primary, secondary, and administrative. In many ways, the identification process at an MFI is a function of integrating the AM and PM information. This chapter will specifically review the principal challenges an investigator faces in reconciling such data.

Identification Errors at an MFI

The remains of victims at disasters are sometimes fragmented, so a victim's primary AM data will often match the primary PM data from multiple sets of partial remains. The reverse, however, is problematic: Primary PM data from a set of remains linking to more than one AM file should undergo further review. Such cases are usually the result of commingling, investigative mistakes, or administrative error.

MFIs are often associated with severe physical traumas (plane crashes, earthquakes, explosions, etc.) that can result in the fragmented remains of one victim becoming embedded in the tissue of another victim. Such cases are not always apparent to the naked eye and may not be recognized during triage at the morgue. The tissue of person A may become fused with the bone of person B such that it appears to be articulated to it. Open wounds can suffer DNA contamination from the blood of other victims. Furthermore, recovery workers may place multiple sets of remains in one body bag. It is standard practice to separate such remains into different cases during mortuary intake (e.g., triage), but this is sometimes missed in the initial chaos of the disaster response. Finally, commingling can sometimes occur *during* the autopsy. This happened on occasion in Thailand after the 2004 Asian Tsunami (hereafter referred to as "Tsunami") when fingers or hands removed for prints were returned to the wrong body bag.

Investigative Mistakes

Investigative mistakes can happen in two ways. First, errors may occur during the analysis of the remains or reference items, resulting in flawed primary data. Second, an analyst may make a fortuitous match between primary AM and PM data due to incomplete test results or, with DNA, make a mistake keying in a kinship equation. A complicating factor is that the analysis and matching activities may be performed in different locations by different personnel.

Administrative Error

Administrative error is where the information is accurate but is in the wrong case. This can happen with AM or PM data, but AM records are more at risk, as victims with similar names can have their case files confused. This potential increases if they share other information such as an address or DNA donor. Victims from the same family are especially vulnerable in this regard. But AM or PM files with nothing in common can be erroneously combined if case numbers are confused or tracking numbers transposed.

Example: Commingling

In one case at WTC, two separate remains with matching DNA profiles were linked to John Smith. The anthropology review showed that the first set of remains was an upper torso with a complete mandible and limbs, while the second set of remains was a complete mandible. The case file for the disarticulated mandible stated that it had been recovered from the chest cavity of the remains to which it was matching by DNA.

Odontologists matched the disarticulated mandible to the AM records of Walter Jones. They also confirmed that the mandible attached to the torso was consistent with the dental chart for John Smith. When the second mandible had been recovered from the chest cavity, a tissue sample was taken for DNA testing. Apparently, the force of the impact affixed tissue from the torso so firmly to the mandible that it appeared to be articulated to it and it was sampled for DNA. As a result, the DNA profile for the loose mandible matched the torso and the personal effects of John Smith. At this point, a tooth was taken from the disarticulated mandible and DNA testing showed it was consistent with the reference samples for Walter Jones. Both identifications were finalized.

Example: Investigative Mistakes

Investigative mistakes usually do not result in false matches because two victims would need to have nearly identical profiles, the analyst would have to make a mistake only at that single point of difference, and the error would have to be such that it transforms the data from the first victim to match the second victim. Given the underlying frequency distributions of DNA, fingerprint, and dental profiles (which are all highly heterogeneous), it is more likely that an investigative error would result in a false exclusion. That said, there are two situations where the potential for a false match increases: DNA testing on degraded remains and DNA matching using partial profiles. These are related issues in that degraded remains often produce partial DNA profiles, but they lead to investigative mistakes in different ways.

DNA Testing on Degraded Remains

DNA samples are often tested in batches, with samples placed into wells on a tray. A well can contaminate an adjacent well on the same tray, and tray-to-tray contamination can occur if robotic pipettes carry trace sample from a well in one tray to the corresponding well in the following tray. Two DNA profiles in one well will be flagged during the technical review, but a well with sample from remains too degraded to yield DNA will only show the profile from the *contaminating* well. In this way, the DNA profile of one set of remains can be incorrectly assigned to a second set of remains.

In one WTC case, DNA linked two remains to the same victim, but an anthropology review showed that they could not be from the same person. The anthropologist noted that one of the remains was badly burned, suggesting its DNA was compromised. This remain was retested, it failed to give a DNA result, and it was removed from the identification. A review of the laboratory process located the source of the contamination and it was fixed. This illustrates the need for a multidisciplinary approach: Anthropology not only caught a mistake, but it alerted the DNA laboratory to an operational problem.

Partial DNA Profiles

A second scenario where an investigative mistake can result in a false match is when working with partial DNA profiles. An error at a single point of comparison will usually not be enough to make the full profiles of two different people appear to be the same, but degraded remains may yield only a handful of genetic markers. When compared to the AM data, it is possible that the difference between this limited PM data set and the complete AM profile is so small that a single error of interpretation may be sufficient to lead an analyst to make a fortuitous match.

Example: Administrative Error

In the spring of 2002, a whole body was recovered from WTC's Ground Zero with personal effects for "Jason Arturo." An arm had already been identified as Arturo's by DNA in the fall of 2001. The whole body was tested for DNA, and it had a different profile than the arm. The laboratory file for Arturo showed that six separate DNA reference collections were received from the FAC in the fall of 2001. But an audit of the collection logs from the FAC showed that only *five* DNA collections had taken place for Arturo (this log had not been available to the laboratory when the arm was identified as Arturo).

At the FAC, victims were given case numbers while each DNA collection received a tracking number. The extra collection number in Arturo's laboratory file was 453. The log at the FAC listed collection 453 as belonging to Nathan Riley, victim number 227. One of Arturo's DNA collection numbers was also 227. The author reviewed the errant package and noted that numbers 453 and 227 were present, but with no label to indicate which was the victim number and which was the collection number. Riley's name was also missing, a failure of the standard operating procedure at the FAC. The log showed that collections 453 for Riley and 227 for Arturo took place on the same day, so there was an overlap in their handling. Based on the review, it was apparent that Riley's victim number had been confused with Arturo's collection number and that his reference samples had gone in the wrong folder. After the error was recognized, Riley's samples were given their own

case file. All of Arturo's reference samples were tested, and they matched the whole body recovered in the spring of 2002.

It has been pointed out to the author that if all the reference items had been tested, then a conflict among Arturo's samples might have been noticed much earlier. While true, this ignores the fact that thousands of reference items were pouring into the laboratory, some of dubious quality. Testing everything would have delayed the identification process by months, without appreciably improving the chances of an identification. In order to spot conflicts within a case (or suspicious matches between unrelated cases), all of the AM testing would have to be completed first. Refusing to release the first DNA match until the last reference item has been tested is probably not a viable policy given the urgency of releasing identifications as soon as possible. However, it *is* possible to compare the FAC collection logs to the case files in the DNA laboratory when a match is made, and an audit of this kind would have flagged Arturo's case. Once the FAC collection log was available to the DNA laboratory at WTC, such audits became a routine part of the admin review.

Likewise, while conflicting PM results from a single set of remains will be discovered during a case review, this assumes that remains will not be released until all primary methods are complete. Again, pressure to release identifications in a timely manner may trump such a policy, and many fragmented remains are simply not amenable to multiple examination techniques. For these reasons, one cannot rely on conflicting primary data to flag commingling, investigative mistakes, or administrative errors. However, an investigator can review the secondary and administrative data for matches, as inconsistencies at this level would suggest that further review is needed. For example, with WTC, the husband of Donna Parker donated his wife's hairbrush as a source of DNA. The profile matched several sets of fragmented remains, including a hand wearing an engagement ring and wedding band with multiple diamond settings. A review of the property report filed by the husband listed only a single gold wedding band. An investigator called the husband and confirmed that Donna did not have a diamond ring and wore only the plain wedding band. Meanwhile, a search of the jewelry descriptions in the AM records found a match to both rings in the file of an unidentified victim. The hand was removed from the Parker identification and placed under separate investigation.

Summary of Identification Errors at an MFI

At an MFI, an investigator should not treat a match between primary AM and PM data as an identification, but as the start of the investigation. The next step is to reconcile all of the AM and PM data (primary, secondary, and administrative). This admin review can catch commingling, investigative mistakes, and administrative errors that would otherwise go undetected. To audit a match, however, an investigator needs to review all of the information related to a case. As it turns out, locating all of the relevant data can be a challenge, especially with the AM files. The next section in this chapter will examine why this is so difficult.

AM Information Intake Dynamics

At an MFI, numerous people will file missing persons reports for the same victim, but will not do so as a single group. The spouse may file a report the day of the disaster, a sibling may go to the FAC the next day, and the parents might fly into town later. It is unrealistic to expect that they will go to the FAC together. Some informants may have to travel a significant distance, while others are nearby and want to file a report immediately. People in different stages of grief may not be ready to visit the FAC at the same time. And some family members may be estranged and prefer to make separate visits.

In addition to family intakes, employers will provide AM data for staff that perish at an MFI, either as victims or as responders. Because the WTC were office buildings, most of the victims were at work when the attack took place, and their employers provided a great deal of information. Several employers played a liaison role, transferring DNA reference samples from families abroad to the FAC. The Fire Department of New York (FDNY) played a similar liaison role on behalf of the firefighters lost at WTC.

This dynamic of staggered intakes was also quite pronounced after the Tsunami. Survivors filed reports in Thailand for missing family members. At the same time, friends and family in the home countries went to their local FACs to do the same, sometimes mistakenly reporting individuals as missing who had survived and were themselves providing AM data in Thailand on other victims.

If an informant is not aware that a file already exists for a victim, the interviewer will create a duplicate file. Even when the informant knows that a case has already been opened, the interviewer may have to start a fresh intake if he does not have access to the existing file, or he may elect to start a new form, recognizing that different informants sometimes provide conflicting information.

In addition to multiple informants, a single informant may initiate multiple intakes. In the time between a disaster and the establishment of the FAC, an informant may file a report with her local police and then again with the police department in the jurisdiction of the disaster. These reports are usually forwarded to the FAC. An informant may call the FAC to open a report before visiting in person. Once there, she may not be able to complete the intake on the first visit as most people do not know their mother's shoe size or the phone number of their brother's dentist. They may need to gain access to their son's apartment to retrieve a toothbrush for DNA testing. Finally, they may be too grief-stricken to finish the interview in one sitting. At WTC, it was not unusual for a case file to contain several intakes from the same informant. After the Tsunami, tourists filed missing persons reports in Thailand and then again in their home country. After Hurricane Katrina, not only did officials collect AM data in the Gulf, but some municipalities where survivors were relocated also collected redundant missing persons reports.

All of the interviews for a victim need to be integrated at some point. But since most intake forms are quite long, they are rarely completed in full. When looking at files that might be for the same victim, a file manager is usually working with data that only partially overlap. This is especially true for updates from returning

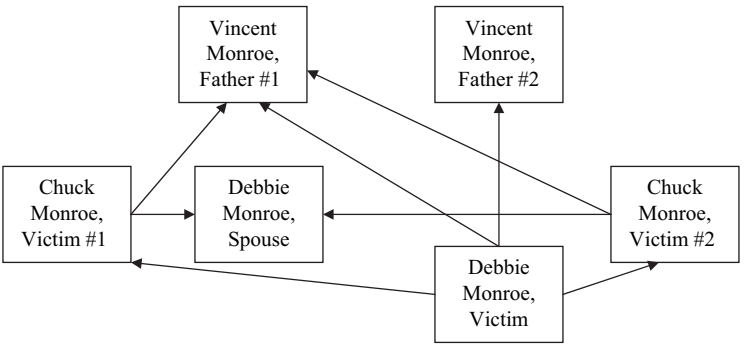

Fig. 18.1 Chuck, Vincent, and Debbie Monroe: one case or three?

informants since the very purpose of a follow-up intake is to provide information that was not available during the previous interview. If different informants give conflicting answers (or if different interviewers record similar answers in a disparate manner), the records manager may hold intakes apart until clarifying information can be gathered. On the other hand, just as comparing partial primary PM data to full AM data can lead to a fortuitous match, so can comparing two partial AM intakes lead to two files being wrongly combined.

Figure 18.1 represents the file from WTC for Chuck Monroe. It contained several intakes from the father, Vincent, and the spouse, Debbie. Not only did the intakes from Debbie and Vincent disagree on some basic issues, but the different intakes from each informant were not internally consistent. Eventually, it was determined that there were *two* victims named Chuck Monroe and that their files had been mistakenly combined. One was unmarried and had a father named Vincent, while the other was married to Debbie. Complicating the matter was the fact that there was also a victim named Debbie Monroe and some of her information wound up in the file for Chuck Monroe because the informant and victim name were transposed. As if that were not enough, the informant for Debbie was her father, a gentleman named Vincent Monroe. At this point it appeared that Debbie and one of the Chuck Monroes might be siblings. Their reference DNA was analyzed to establish that they were unrelated, their files were reorganized, and their identifications went forward.

Case Numbers

Issuing a Missing Person (MP) number for each victim can reduce many problems associated with victim names since it can be used like a case number to manage the files. When returning to the FAC, the MP number can be used to link an update to the correct case, but it is not foolproof:

- An informant may be related to more than one victim and have several MP numbers. If he gets the numbers mixed up, then the data will be linked to the wrong case.

- Different agencies (FEMA, Red Cross, the police, etc.) will issue their own case numbers, which an informant may confuse with the MP number.
- The MP number can be corrupted. At WTC, telephone intakes were given T Numbers (example: T-138) while personal interviews at the FAC got P Numbers (example: P-138). T Numbers were accidentally converted to P Numbers in some databases so that T-138 became P-138. This meant that P-138 was in the system twice, linking to two different victims. As a result, some of the DNA reference samples and AM intakes were misfiled. Eventually, a Reported Missing (RM) number replaced both the P and T Numbers and the errant cases were reconciled.
- Reading MP numbers is error-prone. MP 1221 may get confused with MP 1212. If the victim list is alphabetized and assigned consecutive MP numbers, then victims with the same last names will have nearly identical case numbers: Lee (133), Lee (134), Lee (135). Barcodes can make case numbers more reliable but cannot eliminate these problems. Their initial assignment can be incorrect, subsequent intervention in the tracking database can introduce errors, and incompatible barcode systems in a multi-agency response is almost a guarantee. Finally, even with barcodes, some file handling will still be conducted manually.

Summary of AM Intake Dynamics

Multiple informants per victim and multiple intakes per informant can lead to AM files that are consolidated incorrectly. As a result, a victim's information may be spread out over several unlinked files, each lacking the data necessary to make an identification. And if one file does contain primary data that match a PM case, an investigator may not realize that crucial information exists in another folder. Conversely, if data for victim A are placed in the folder for victim B, then a PM match to victim A will be identified as victim B. To better understand how these two outcomes occur and what an investigator can do in response, we need to examine the AM interview process in detail.

Challenges in the AM Interview Process

There are at least six possible points of failure in the AM interview process:

1. Incorrect information from the informant
2. Informant misunderstands the question
3. Lack of standardized answers
4. Typos and transpositions
5. Handwriting
6. Data in the wrong field

Incorrect Information from the Informant

It is possible that the informant will provide incorrect information. In fraud cases, this is intentional. Once in the system, such information (and reference items) can become mixed with legitimate cases. For this reason, even fraudulent case files must be audited.

At WTC, one fraud case had the same name as a real victim and a personal effect from the legitimate case was accidentally filed in the fraudulent case folder. Even though no DNA testing was scheduled for the fraud case, its file was audited. The errant personal effect was discovered and used in the identification of the legitimate case.

Informants will also unintentionally provide incorrect information. Given the variety of people who go to the FAC (parents, siblings, spouse, children, co-workers, employers, etc.) and the range of questions asked, it is to be expected that a given informant may be mistaken in some of her answers. This results in intakes with conflicting information that an investigator needs to reconcile. Because names and dates are often used to manage case files, they merit special consideration.

- Anglicized names: Foreign nationals often adopt Anglicized names while living in English-speaking countries. Friends, co-workers, and employers may refer to the adopted name while parents and siblings use the birth name. Won Park went by John Park at work, and his AM records at WTC included both names.
- Nicknames: An informant may provide the victim's nickname, such as Katie or Jack, while others provide the full name Catherine or John.
- Alias: Some people use an alias or go by their middle name. "Carl James Anderson" might be reported as such, while others refer to him as "Jim Anderson."
- Incomplete surnames: If a surname has multiple terms, one informant may give just the last term while others provide the full surname. This happened regularly at WTC with Hispanic, Arabic, Asian, and African names. For instance, AM intakes for Hector Tirado Jimenz included both Hector Tirado Jimenz and Hector Jimenz.
- Maiden names: A woman may retain her maiden name at work for business reasons. The employer, co-workers, and old friends may report her missing by the maiden name while immediate family members will provide the married name.

Some of these examples may appear trivial at first. After all, anyone looking at the files for Catherine Bigsby and Katie Bigsby can see the likely connection. But this relies on some cultural familiarity. An American might know that Jack is a common nickname for John, but is that widely known abroad? The popular Southern nickname "Bubba" can substitute for any first name, a fact that may not be widely known abroad. Conversely, how many Americans know that the Russian diminutive for Alexander is Sasha? Or that Jay can be short for Sanjay?

Spotting the similarities between the names on two case files also relies on their being in close proximity to each other. But when the victim list at WTC was sorted by surname, Carl Anderson and Jim Anderson were separated by several entries

(Clyde Anderson, Dave Anderson, etc.), diluting the apparent connection. And a search for "Tirado Jimenez" failed to retrieve this victim's folder, which was filed under "Jimenez." Only after the victim list was sorted by date of birth did these records cluster together.

Another problem can occur because adult children don't always know the year of their parent's birth and may be too embarrassed to admit it, so they estimate. And couples can marry privately and have a public wedding later. In such cases, the wedding date provided by the parents will not match the date given by the spouse. This is important when trying to link the date inscription on a ring to an AM file.

Informant Misunderstands the Question

"What is your relationship to the victim?" An informant at WTC answered "Father," but his DNA profile did not fit the family pedigree. A review of the intake showed that his date of birth was after the victim's, which meant that the victim was *his* father. The pedigree was revised and it worked. This shows how administrative data (the donor's date of birth) can resolve problems. Other questions that can be misinterpreted by the informant include "Address?" (the informant's or the victim's?), "Parents of the victim?" (biological or legal?), and "Where did they work?" (location or employer?).

Lack of Standardized Answers

If interviewers use different terms to record the same information, then a search of the AM database by one term will only find a portion of the relevant records. And if the AM and PM nomenclature is not consistent, then searching either database for investigational leads becomes problematic. Most intakes use checkbox answer menus where possible, but some concepts are not so easily categorized and the interviewer has to manually record the data, leading to more nonstandardized answers.

Jewelry descriptions in the Tsunami database included "YM," "yellow metal," and "gold." All are accurate, but a search by one term will fail to find the records using the other term. A PM report at WTC for the top half of a head read "male, hair = none." A search of the AM database for men where hair was "none" failed to return the records for bald men. While all manually recorded data are subject to this problem, names, dates, and object descriptions are particularly vulnerable.

Names

There are three main standardization issues with regards to names: generational identifiers, multi-termed names, and word order.

Generational identifiers such as Junior, Jnr., and Jr. are all accurate, but if interviewers use all three, then separate intakes for the same victim will often lack matching names. A search for "Smith, Jr." will not find "Smith, Junior," and vice versa.

Multi-termed surnames present two challenges. The first is determining where the surname begins. The second is deciding how to handle hyphens. In a case from WTC, a woman named Alice Mayhew-Jones had separate AM intakes filed under

- Alice Mayhew-Jones
- Alice Mayhew Jones
- Alice Jones
- Alice Mayhew (filed by the parents, who were estranged from their son-in-law)

Sorting the records by last name meant that Mayhew-Jones and Jones did not cluster together. Searching by "Jones" or "Mayhew" retrieved only half the records. And searches by "Mayhew Jones" failed to return Mayhew-Jones.

Word order: In English, it is typical to address people by their forename and list names alphabetically by surname. But some cultures do the opposite. If an American interviewer is told that the victim's name is "Kim Won Duk," he may reverse the terms to correspond to the "surname, forename" format, as in "Won Duk, Kim." But if the informant gave the name in the Korean style (surname first), then Kim is the surname, not Won Duk. Other informant-interviewer pairs may generate different combinations for the same victim.

Dates

The European format is Day/Month/Year while in the United States it is Month/Day/Year. Since many disasters in the United States have an international dimension, both date formats will likely appear in the AM records, especially if foreign employers, embassies, or police agencies are involved in collecting information from the home countries. A wedding date on a foreign-sourced intake may read 7/3/75 while a PM report on wedding band from a set of remains may be recorded as 3/7/75. Complicating the matter is that some common software programs format numeric dates as text. In this way, it is not immediately apparent that March 7, 1975, and July 3 could be the same date.

Objects

Certain objects associated with a victim can be useful to an investigator, such as jewelry, clothing, and tattoos. Unfortunately, it is very difficult to standardize their descriptions, as the "YM vs. yellow metal vs. gold" jewelry example illustrates. For the Tsunami, AM reports came in from around the world, and there was a multinational presence in the morgue. As a result, AM and PM descriptions often

reflected national variations in describing the same object. For example, swimwear was alternatively described as swimsuits, bathing suits, bathing costumes, bikinis, swim trunks, or Speedos.

A final example illustrates how widespread nonstandard entries can be. At WTC, searching the AM database for employees of "Cantor-Fitzgerald" would not retrieve records for "Cantor Fitzgerald" (no hyphen), "Cantor Fitz," or "Cantor." Each entry is accurate and clear to anyone reading it, but these variations made it challenging to find all of the cases related to the company.

Typographical Errors and Transpositions

All manual intake systems are prone to typographical errors, and MFI data are no exception. For example, in the clothing descriptions of the AM files at WTC, there were 26 variations on the spelling of "khaki." Transpositions can occur in phone numbers, addresses, and birth dates. Even names can be transposed. An intake for "Andrew Clarke" was filed under "Clarke Andrew." In more than a few cases, the informant and victim names were transposed. Not only did this make it harder to find the AM data for a victim, but it put a living person on the list of the reported missing.

Handwriting

Many intake systems are handwritten. Not only is handwriting hard to read, but it encourages the use of abbreviations and acronyms, which may not be obvious to the reader. These forms often provide only limited space to record answers, forcing one to write smaller (reducing legibility) and use more abbreviations. When later entered into a database, there is the possibility of transcription errors. For example, the number 4 can look like a 9, the letter a can be confused with the letter o and so on. The victim list at WTC had an entry for AHOV, but no AM intake could be found for this person. Eventually, the file was found under the correct name, Attov. The two lowercase ts had been read as an uppercase H.

Data in the Wrong Field

If an AM tattoo description is logged in the field for clothing descriptions, it is effectively hidden from an investigator searching for a match to a tattoo on a set of remains. A more common problem is that data can often be recorded in more than one place. For example, many forms have fields for describing a wallet and its contents. Most intakes also have a section for the driver's license. Because these are often found in the wallet, it is not unusual to see them documented in the wallet section instead.

The generic instance of this dynamic is the intake question with checkbox answers where the last choice is "Other," accompanied by a blank text field. While this design has obvious merit, it means that the answer may reside in the text field or the checkbox. In addition, most forms have a section at the end for "Other Information." Data are often entered here despite the existence of a dedicated section elsewhere on the form. As a result, searching the answer checkboxes to a dedicated question will often miss relevant records.

Error Rates in the AM Records

The Disaster Mortuary Operational Response Team (DMORT) is an agency of the U.S. government that assists local medical examiners at MFIs. They have an AM/PM data collection system called the Victim Identification Program (VIP); the AM form is 7 pages long with nearly 300 data fields for about 200 questions. There are more fields than questions because some questions ask for several pieces of information, such as addresses (street, city, state, and zip code). The author's experience in working with VIP at WTC was that an average intake completed about 200 data fields.

If there are six variables in capturing information correctly (detailed above), then we can estimate the accuracy of an intake with the formula (%^N), where % is the chance that a variable is executed correctly and N is the number of variables. This assumes the variables are independent and that the chance of success for each is constant. If an AM interview is 99% accurate at each of the 6 possible points of failure, then an answer is estimated to be correct 94% of the time (.99^6). The estimated error rate is 1 minus the estimated accuracy rate, or $1 - 0.94 = 0.06$. With 200 data fields, a 6% error rate yields an estimated 12 errors per interview.

But 99% accuracy may be too much to expect. Consider sitting in the comfort of your home buying books on the Internet. You are providing a small amount of personal information that is well known to yourself. There is no handwriting and no interviewer, just you entering data into the computer. Yet, it is not uncommon to make mistakes in this setting, suggesting that the accuracy rate may realistically fall below 99%. Now imagine you are at an FAC in a distant city being interviewed by a complete stranger regarding a loved one presumed dead in a disaster. The interviewer asks numerous personal questions, such as your wife's bra size or whether your brother was circumcised. Was he ever arrested? Who is his dentist? The interviewer is polite but hastily writes down your answers. Later they will be entered into a computer by someone else who has already read dozens of intakes that day.

In this context, 99% accuracy might be optimistic. If we lower it to a more realistic 95%, then the estimated accuracy of a data field is (.95^6) = 75%, or an error rate of 25%. For 200 data fields, this is 50 problems per intake. A plane crash with 150 victims and 3 informants per victim would have about 22,500 data problems ($150 \times 3 \times 200 \times 0.25$). If each informant has a follow-up interview during which

she completes another 50 data fields ($150 \times 3 \times 50 \times 0.25$), that is another 5,625 problems, for an estimated total of 28,125.

It is worth stressing here that this does not mean the data are wrong (although that is possible). Problem data in this context can mean that a correct answer is in the wrong field or was recorded in an irregular manner (Jr. vs. Junior, hyphens in the surname, European date format, etc.). It may have a minor typo or transposition or answers a different question than the interviewer intended. As a result, the record will not be found by a search or sort of the files using these criteria. Linking updates can also be challenged by such problem data.

Assume that there is one informant per victim and each makes an initial visit to the FAC, creating a case file. Each informant then returns a few days later to provide a DNA sample. With 150 victims, that is 150 updates that have to be linked to the correct case. If this process is 95% accurate, then 5% of the follow-ups are at risk. If these 7 or 8 updates are added to the wrong folder, then about 15 cases would be impacted, or 10% of the total. One can see how this number increases with each successive round of updates and if more than one informant is involved.

At WTC, there was a victim named Henry Niles and another named Henry Niles, Jr. They were unrelated. The son of Henry Niles made several trips to the FAC, as did many members of the family, resulting in numerous intakes and case numbers. The son was named Henry Niles, Jr., and on one visit to the FAC his update was accidentally placed into the file for the victim named Henry Niles, Jr., because the person handling the file confused the informant name with the victim name. Unfortunately, this update contained DNA reference samples. Eventually, these profiles linked to a set of remains that were incorrectly identified as Henry Niles, Jr. The error was only discovered during the admin review.

Error Rates in the PM Records

PM records are vulnerable to (1) lack of standardization in recording data, (2) handwriting issues, (3) typographical errors, and (4) storing data in the wrong field. Assuming 99% accuracy at each possible point of failure means a record is 96% accurate (.99^4). PM forms often mirror the AM intakes in organization and can run several pages with hundreds of data fields. Not every topic will be relevant for each set of remains, so assume that only 200 data fields will be completed. That is about eight errors per record.

Cases may be reexamined and these updates need to be linked to the correct folder. If 25% of all autopsies are reexamined and 5% of these are incorrectly linked, then there will be about a dozen update errors per 1,000 cases. Since each error affects 2 cases, about 2.5% of all PM cases will have a filing error ($1,000 \times 0.25 \times 0.05 \times 2$).

For example, assume there is an MFI with 200 victims and each body fragments into 7 pieces (at WTC, nearly 20,000 remains were recovered for about 2,750 victims). That is 1,400 PM cases. Eight mistakes per intake and a 2.5% updating error

rate mean a PM database with 11,235 problems. Again, this does not mean just errors of fact, but correct data that are recorded in a manner that makes it difficult for an investigator to pursue a match or consolidate records.

Addressing the Problems

While an ounce of prevention is certainly worth a pound of cure, the nature of disaster response seems to guarantee that there will always be problem data. This section will look at prevention first and then address how to work with problem data.

Preventing Problems

There are several steps an agency can take to reduce the data errors at an MFI. First, it should define the AM/PM intake forms it will use ahead of time. Since local authorities will likely receive assistance from DMORT, it is a good idea to preview VIP, available at www.dmort.org. Interpol also has a series of Disaster Victim Identification intake forms that may be helpful (www.interpol.org). An agency could use one of these systems and create supplemental forms to reflect local needs, or it could design its own MFI system based on the missing persons forms already used in its jurisdiction.

Once the intake system has been defined, standardized recording formats should be adopted for names (generational identifiers, multiple-termed surnames, hyphens, etc.) and dates (MM/DD/YYYY or DD/MM/YYYY). The rest of the questions should be reviewed to standardize answers where possible, particularly the abbreviations and acronym for describing objects, concepts, and affiliations. It is critical that both PM and AM intakes use the same nomenclature. This is not a matter of finding the "correct" format so much as having a consistent approach.

Rather than limiting this process to those who will collect the data (mortuary staff and interviewers at the FAC), it should also include the database managers who will administer the system as well as the investigators who will use it to make identifications: medical examiners, pathologists, odontologists, anthropologists, fingerprint examiners, and DNA analysts. DNA is noteworthy here because not only are DNA experts often left out of disaster preparedness planning, but they are sometimes not even included in the immediate disaster response because of the perception that DNA is not timely.

When finally deployed, DNA often inherits policies and systems that do not reflect the needs of forensic biology, impeding their effectiveness. Concerns regarding the pace of DNA testing are often deflected by calling it the method of last resort, and this is dutifully repeated by the media, perpetuating the notion that DNA methods are slow. In many ways it is a self-fulfilling prophecy. A notable exception to this mindset is the close coordination between the Armed Forces DNA Identification Laboratory (AFDIL) and DMORT. Not only are the protocols in place for DNA

testing prior to a disaster, DNA teams from AFDIL deploy with DMORT as part of the initial disaster response.

Computer resources should be secured to avoid handwritten intakes. Staffing rotations should be established for the morgue and the FAC. Liaisons at partner agencies should be included in the planning stages and vendors expected to provide services, equipment and consumables should likewise be included to ensure that they have the necessary surge capacity.

Conduct mock AM interviews. Not only is it valuable training, but it will test the efficiency of the intake system and may help flag individuals who really are not suited to interviewing grieving families.

Barcode systems should be established and checked for intra-agency compatibility. Tracking labels based on the barcode numbers should be readable by humans, and a case name/numbering schema should be developed to create a unique identifier for every object in the system. On the PM side, this would include body bags, recovered property, and samples taken for fingerprinting, odontology, or DNA. On the AM side, there are informants, victims, intakes at the FAC, and reference materials submitted for fingerprints, X-rays, and DNA.

There are four general considerations in developing the naming schema. First, it has to contain enough information to allow a human operator to work with the object. A person should be able to tell at a glance if he is holding an AM or PM folder.

Second, toe tags, file folder tabs, X-rays, test tubes, etc. can be quite small. This places a practical limit on the length of the label name. Using a tiny font to make a name fit does not support the first guideline. In these cases it may be prudent to place small items in larger containers that can support larger labels.

Third, the label should not include any information that may be modified or updated. As has been shown, victim data can change with each intake, even for seemingly static concepts like name and date of birth. The problem is that corrections to a label are not backwards-compatible. The old label already exists on boxes, on sample tubes, and in printouts. It resides in spreadsheets and databases that may not be available for updating. An investigator working with a changed label will not be able to trace a reference sample backwards through the system to intake, voiding its chain of custody.

On the other hand, uncorrected label names are a source of confusion. At WTC, a reference sample for Jonas Clare had the prefix BM for Biological Mother. But it had a male profile. A review of the AM intakes for Clare showed that the spouse brought their infant son to the FAC to provide DNA. Mrs. Clare signed a consent form for minors and under "relationship to donor," she wrote "mother." The interviewer entered this as the relationship to the *victim,* and a tracking label was created with the prefix BM. This sample was flagged as a problem by every new analyst who saw it, consuming countless hours of staff time explaining the situation or reinvestigating the matter. But changing it to BS (Biological Son) would have created a different problem: The collection log for Clare at the FAC (maintained by a separate agency) still showed a single DNA donation from the mother and all of the transaction records for this case listed a BM sample. An investigator starting

with the BS sample would not be able to work backwards to find a DNA collection from the son of Jonas Clare, and the chain of custody would have been broken for a perfectly viable sample. In the end, it was decided that being able to track the sample through the identification process was more important than a name with the correct relationship code, a decision the author wholly supported. From a data management point of view, it is better to have a name that is flawed but consistent than one that is inconsistent. But the best approach is simply not to embed data about the sample in the sample name, as this information can and will change.

The fourth guideline for creating a naming schema is that it be compatible with systems already in place. Collaborating agencies and vendors will have their own information management systems, complete with accessioning numbers, data fields, reporting tools, and standard operating procedures. The more information one tries to include in the label name, the more likely it is to conflict with these preexisting systems. For instance, one could mandate that label names for property include the letter R if it is a ring. The evidence division at a supporting police agency may already have a protocol of using J for jewelry, followed by a number to indicate the type: 1 = ring, 2 = necklace, etc. Everyone working on property will have to translate R to J1 and back again. And they will have to perform separate translations for each property type. The situation becomes even more cumbersome when the label name attempts to capture information from the internal process at the supporting agency. DNA testing is a complex undertaking with several process steps and the potential for samples to repeat certain steps. While the operating parameters in place at each step may be useful to the analyst interpreting DNA results, there is no need to squeeze this into the sample label. Such data are fully available in the laboratory's information management system (LIMS) and can be linked to the sample name as a supporting file.

Ultimately, labels designed to display information regarding the handling of an object are easily compromised. In Thailand after the Tsunami, body bag numbers were based on the nationality of the autopsy team. Unfortunately, countries rotated staff every few weeks, reducing the specificity of this number. Other autopsy teams were multinational in composition, rendering the number ambiguous. As bodies were reexamined, autopsy teams added new body bag numbers to reflect their nationality. This became a significant source of confusion as many body bags eventually had more than one tracking number.

Solving Problems

In order to audit a proposed identification or create an investigational lead, one must first assemble all of the potentially relevant AM records. First, conduct multiple searches that are broad enough to ensure that no relevant record is missed (this approach accepts that irrelevant files will be also retrieved). Second, compile the searches and cull unrelated cases by sorting the list via a demographic detail and removing the records in clusters that are disqualified. Manually review the

remaining cases to consolidate or exclude. For example, search for Employer = Acme and Gender = Male. If the victim lived in New Jersey, sort the results by State and omit all entries except NJ.

Some basic tips for conducting broad searches:

- Use truncated surnames with wildcards to overcome typographical errors: Rama* will find Ramachandran and Ramashandran.
- Phonetic variations can find misspelled names: Lee & Li, Cane & Kane.
- Wildcards can find nonstandard uses of generational identifiers and hyphens: Hernan* will find Hernandez-Tirado, Hernandez Tirado, Sr., and Hernandez, Senior.
- Objects can be described in a variety of ways, so try synonyms like gold & yellow, brown & brunette, or bikini & swimsuit.
- Consider acronyms and abbreviations, such as YM vs. yellow metal or FDNY vs. Fire Department, New York City.
- Try reverse searches: If you are looking for the files on Terry John, try putting Terry in the surname field. If a ring has a date inscription of 11/8/83, query the wedding date fields for that as well as 8/11/83.
- Search related fields such as (wallet/purse and driver's license). Query catch-all summary fields in associated sections of the intake.
- Do not use AND with multiple field searches. Remember that many fields in an intake will be left blank and records lacking data in a field will be excluded by searches that mandate data be present. A search for "Asian" AND "Male" AND "Tattoos = Yes" will not return records for tattooed males where race was left blank.

Again, the key to finding all the potentially relevant records is not to conduct a single, precise search that finds "the" record for a case. In fact, it could be argued that the only thing worse than a search that returns no matches is a search that returns some matches, creating a false sense of success. The search strategy has to exhaust all possible avenues of linking to a relevant file.

The heuristic of conducting several broad searches and then excluding irrelevant records is a tactical approach for researching individual cases. But there is also a need for a comprehensive, systemwide view of the data. First, one should map the AM and PM operations in order to identify the links in the chain where material and/or information is handed off. Once the process has been mapped, then the inventory manifests between each link can be compared to reveal missing or surplus items.

For example, remains will be recovered at the disaster site, moved to a local collection area, transported to the morgue's holding facility, pulled for autopsy, and placed in storage. A different group will carry out each step, the performance of which will generate time and date stamps, employee names, and details specific to the function at hand. Search teams will note the location where remains were found, evidence techs will list the property recovered, pathologists will provide autopsy findings, and mortuary attendants will note where the remains are stored. In addition to the body bag number and a master PM file that travels with the remains, each

group may have its own recordkeeping system complete with tracking numbers, barcodes, and databases.

If the morgue accessions a dozen set of remains from the disaster site in a single day and all are autopsied, then mortuary storage should also accession a dozen set of remains that day. If not, that would be a cause for review. At WTC, the accession and inventory numbers did not always match, because as anthropologists reviewed the remains in autopsy, they discovered cases of commingling. The remains would be separated, creating a new postmortem case. In other instances, they determined that the remains were nonhuman and removed them from the system. Both were sound policies and accounted for the disparate numbers at different points in the process.

A critical component of mapping the system and reconciling inventories is defining the terminology used at each link in the chain. The FAC at WTC generated a P Number for each victim and a DNA Case Number for each DNA collection. This information, along with the DNA references, was sent to the New York State Police DNA laboratory. The NYSP database already had dedicated fields for tracking data, so the P Number was filed under their Incident Number and the DNA Case Number went in their Reference field. The Case field in the NYSP database was then populated with a tracking number called the WDI Number. When tracing DNA samples through the system, it was important to clarify which Case Number was being referred to, the one from the FAC or that from the NYSP. It was also critical to know that the P Number at the FAC became the Incident Number at the NYSP and that the DNA Case Number became the Reference Number.

Generic terms like case, file, collection, incident, sample, reference, and item were routinely used by agencies at WTC with different meanings, a phenomenon the author witnessed again in the Tsunami response. While it may not be possible to standardize the usage of these terms across all links in the chain, defining their different meanings can reduce handling errors and facilitate tracing an item's chain of custody.

Summary

Preparing for the information management needs of a disaster ahead of time may not eliminate all of the data management problems that can arise when numerous agencies collaborate on a project, but it can reduce the scope of the problems that will emerge. A sound data management strategy after the disaster will allow for the effective integration of AM and PM records in support of the identification process. The most critical dimension of this process is using the data to create investigational leads and reconcile all of the information (primary, secondary, and administrative) when there is a match between AM and PM data.