



The author(s) shown below used Federal funding provided by the U.S. Department of Justice to prepare the following resource:

Document Title: Study of Familial DNA Searching Policies and Practices: Cost Simulation Tool User Guide

Author(s): Avi Bhati, Sara Debus-Sherrill

Document Number: 251045

Date Received: August 2017

Award Number: 2013-R2-CX-0013

This resource has not been published by the U.S. Department of Justice. This resource is being made publically available through the Office of Justice Programs' National Criminal Justice Reference Service.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Study of Familial DNA Searching Policies and Practices

Cost Simulation Tool User Guide



Avi Bhati and Sara Debus-Sherrill

June 2017

This document describes a tool that can be used to estimate or simulate expected costs associated with adopting familial DNA searching (FDS) technology. The FDS Cost Simulation Tool uses two types of information to estimate potential costs: (1) the probability that a FDS request will progress through various stages of the process and (2) costs associated with each stage during lab processing. The information was gathered from study interviewees and expert advisors from the Study of Familial DNA Searching (see *Methodology* box at the end of this document for more information about the study).

The FDS Cost Simulation Tool has six tabs, including: (1) an “Introduction” tab which provides some overview information about the tool, (2) the “TransitionsData” tab where users can enter information about DNA case processing, (3) the “TransitionsDataSample” tab where users can view example data from the study, (4) the “CostData” tab where users can enter information about expected costs associated with various stages of the familial DNA searching process, (5) the “CostDataSample” tab where users can view example data from the study, and (6) the “FDSCostSimulationTool_Summary” which provides a summary of estimated costs based on the information entered. As a default, the example study data is already included in the “TransitionsData” and “CostData” tabs, but users can overwrite this information if they have data more specific to their own jurisdiction/community. The rest of the tool has protected cells to avoid overwriting formulas and to preserve the example pilot data from the study. This document details how to enter and interpret the information in these various tabs.

Input 1: Stage Transitions

The tool lists five key stages that an FDS request may typically pass through during lab processing.

1. *FDS Request* – To initiate use of FDS as an investigative tool, an agency must first request permission for the crime lab to perform FDS. Permission may be denied or approved.
2. *FDS Software Search* – If approved, the next stage is a search using specialized FDS software. The search may result in a potential familial association above a certain threshold likelihood level or may result in no associations identified at or above that level.
3. *Lineage Testing* – If a potential association is found with a male relative, the crime lab will next conduct a lineage test (Y-STR) to further support relatedness. If the potential association is female, the crime lab may perform lineage testing. However, few labs have this capability currently, so instead those cases may proceed to either an internal review of other non-DNA information (e.g., birth/death records) or to the direct release of the identity to law enforcement.
4. *Review of Non-DNA Information* – The lab or state DOJ may next conduct a review of additional records information to help build family trees or narrow down suspects (e.g., if there are two brothers but one died before the crime occurred). This stage sometimes occurs prior to release of information to law enforcement and other times is performed by law enforcement themselves.
5. *Release of Identity* – After all these stages, if the potential familial association still appears to be a viable lead, the identity of the possible family member may be released to law enforcement. If FDS yields a potential association with a female family member, her identity may be released to law enforcement directly without a lineage test or an internal non-DNA review.

For more information about this brief or study, contact abhathi@maxarth.com or Sara.Debus-Sherrill@icf.com.



This project was supported by Award No. 2013-R2-CX-0013, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the U.S. Department of Justice or the roundtable participants.

All FDS cases begin at the FDS request stage (A) and then pass through one or more of the subsequent stages (B through E). A given case does not need to pass through all stages and may drop out of consideration at any of the subsequent stages or may skip certain stages. For example, a potential male family association which is not confirmed by lineage testing will not move to subsequent stages. Similarly, a database search that yields a match on a female familial relationship will typically not undergo lineage testing, but the identity of the potential female family member may be released to law enforcement. Many labs do not perform their own review of non-DNA information and instead leave this to law enforcement to perform.

Entering Transition Data in the Tool

In the FDS Cost Simulation Tool, on the tab labeled “TransitionsData”, the cells in Table D are used to compute the likelihood of transitioning through the various stages of the FDS process. The column labeled “Transition Counts” (Column G) provides a summary of data obtained from two states who provided pilot data. States and communities interested in using the tool to anticipate potential costs may use this same data for their own estimates (assuming that the probabilities associated with each stage are relatively comparable across jurisdictions), which is pre-populated in the cells. Please note that this data is based on the experiences of only two states (two state-level labs and one local-level lab within one of these same states). Every jurisdiction will be different, and this may influence the accuracy of final estimates.

Alternatively, if jurisdictions have their own data on the likelihood of progressing through these FDS stages or have different estimates they would like to use, they may overwrite this data with their own stage transition probability data in the green cells under the column labeled “# of Cases in Your Jurisdiction” (Column E). Transition rates are automatically computed under columns H and I, to be applied to the cost data. The tab, “TransitionsDataSample,” preserves the pre-populated pilot data and may be used as an example even if users overwrite the transition data in the “TransitionsData” tab.

Input 2: Cost Data

Cost data can be broken down into several categories of fixed (investment/start-up) and variable (per case) costs. The costs should be computed as marginal (over and above those associated with a traditional DNA exact-matching case). The FDS Cost Simulation Tool includes the following categories of costs:

Fixed/Investment/Start-up costs:

- *Equipment/Hardware/Software* – Costs associated with any physical assets (computers, servers, printers, or other hardware) the lab or law enforcement needs to acquire in order to implement FDS. This also includes costs associated with developing, validating, or otherwise testing the software or process.
- *Training* – Costs associated with developing training materials and to conduct initial training.
- *Other* – General cost category relating to any other costs associated with implementing FDS, such as resources required to enact and obtain approval for new policies or to hire additional staff for supporting the FDS functions.

Variable costs:

- *Personnel* – Labor hours spent to prepare, submit, and process FDS request; conduct lineage testing; perform non-DNA information review; or conduct/receive ongoing training.
- *Supplies* – Any supplies related to storing, transporting, and backing up FDS results/findings; reagents, plate, and other lab/chemical supplies; or supplies related to non-DNA review efforts.
- *Travel* – Travel related to attending informational meetings about FDS or to receive FDS results; gathering information for conducting non-DNA information review; or any other travel required.
- *Software Fees* – Ongoing fees may be needed depending on licensing arrangement for conducting FDS.



- *Contracted Services* – Costs associated with outsourcing any part of the FDS process (e.g., lineage testing), if applicable.
- *Other* – All other variable costs

For the ongoing/variable costs, the cost data are structured as per unit (cost) and number of units (multiplier) data. For example, personnel costs are broken down by the hourly wage rate and number of hours. Similarly, travel costs can be broken down into \$/mile and number of miles. A jurisdiction may also enter a fixed dollar amount by entering a single unit (e.g., a wage rate of \$50 per hour utilized for 10 hours may also be entered as a wage rate of \$500 utilized for 1 hour). For varying wage levels within a particular stage, it is advised to use the average wage rate (i.e., $[(\#hrs \text{ at wage rate } 1 * \text{ wage rate } 1) + (\#hrs \text{ at wage rate } 2 * \text{ wage rate } 2) \dots \text{etc.}] / \text{total } \# \text{ hrs across wage rates}$). While data may be entered for every category, that may not be required for every jurisdiction. For example, if a jurisdiction does not perform reviews of non-DNA information or does not incur ongoing software fees, it can leave these categories blank, and the data will be treated as 0.

Some labs may use bulk processing for lineage testing instead of individual testing for each sample/profile comparison. The cost simulation tool also provides the ability to estimate costs for bulk processing if the lab uses this approach.

Entering Cost Data in the Tool

Similar to the “TransitionsData” tab, jurisdictions can either use the existing sample/pilot data for expected costs or enter their own cost information. Cost data are divided into three tables—(A) Fixed costs, (B) Variable costs for individual FDS, and (C) Variable costs for bulk FDS. Sample cost data for each of the tables is provided by one local-level crime lab in the tab labeled “CostDataSample.” Similar to the caution noted above for stage transitions, this cost data reflects the experience of this one crime lab, and different labs may have different expenses. Jurisdictions may overwrite the green cells in the “CostData” tab with their own cost estimates for each stage of the process. For variable costs involving bulk or batched FDS processing, the user also needs to specify the number of cases that are included in a typical batch. This information is entered in the last row of Table D on the “CostData” tab. In the example provided, this number is 50, but this may vary depending on individual lab practices.

Interpreting Overall Costs

The last tab in the tool, “FDSCostSimulationTool_Summary” provides an overview of the cost model. Once all available data has been entered into the tool, a jurisdiction will obtain summary estimates of the expected costs associated with FDS. The tab first provides information on the total fixed or investment/start-up costs associated with adopting FDS technology. Start-up costs are fixed and incurred irrespective of whether a jurisdiction utilized individual and/or bulk requests. Beyond the start-up costs, the tab next presents estimates of the probability that an FDS request will result in the identity of a potential familial relationship being released to law enforcement. This may happen under two different scenarios—(i) the identity of a male familial relationship may be released after being confirmed via lineage testing or (ii) the identity of a female familial relationship may be released to law enforcement without lineage testing. The former will result in fewer identities being released to law enforcement but with higher chances of resulting in a viable lead. The latter will result in many more identities being released to law enforcement but with lower chances of resulting in a viable lead. We have included summary cost information under both scenarios since either (or both) may occur depending on the specific case and jurisdiction policy. All results are therefore provided in two columns—labeled to indicate the scenario being considered.

The sample data from two states suggest that the probability of a potential family association being released to law enforcement authorities is 7.7% when only IDs supported by lineage testing are included, but as high as 40% when all released IDs are included (including non-confirmed female associations). The sample cost data from one local-level crime lab demonstrates potential start-up costs of approximately \$23,000 and suggest that the expected marginal cost of following through with an individual FDS request is \$738 (cell D12 on the tool’s last tab) when including only lineage tested cases

and \$771 (cell E12) when including all released IDs (including female associations).¹ In a similar manner, the expected marginal cost of following through with an FDS request being bulk processed is \$211 (cell D13) when including only lineage tested cases and \$214 (cell E13) when including all released IDs.

While the above cost computations provide marginal costs associated with each FDS request, it is also interesting to estimate the expected price for each potential family association released to law enforcement. The sample data suggest that it would cost up to \$1,910 (cell E16) for every ID released to law enforcement using individual FDS requests, while it would cost up to \$530 (cell E17) for every ID released to law enforcement using bulk FDS requests. However, should we be interested in the higher quality leads—only potential family associations supported by lineage tests—then the costs are different. The sample data suggest that, for ID releases that are further supported by lineage testing, it would cost up to \$9,600 (cell D16) for each ID released to law enforcement using individual FDS requests while it would cost up to \$2,741 (cell D17) for each ID released to law enforcement using bulk FDS requests.

The cost summary table will automatically update if a jurisdiction provides their own data in the cost and transition tables (only cells shaded green need to be modified). Note that by inputting different values for the transition counts as well as cost numbers, one may simulate a number of hypothetical scenarios and use the tool for decision-making. Related to the pre-populated transition and cost data shared in the tool, it is important to remember that these are sample estimates from a small number of jurisdictions and should be **interpreted with caution** due to the small sample size and variability in experiences across jurisdictions.

METHODOLOGY

ICF, with support from the National Institute of Justice, conducted a multi-phase study on FDS policies and practices in the United States. This study had multiple components, including two expert roundtables, a systematic literature review, a policy scan of practice, a national survey of CODIS laboratories, a series of state case studies, and cost modeling. This study was intended to fill knowledge gaps and provide information about this emerging practice.

Data on general cost factors were collected from project advisors and case study interviewees, while detailed cost estimates were collected from a local-level lab who agreed to provide these estimates to pilot the FDS Cost Simulation Tool. Interviewees had mixed opinions on whether the costs of investigation were different for FDS cases compared to other types of DNA cases; therefore, this tool focuses on costs prior to investigation by law enforcement once they have received the FDS investigative lead. Data on the number of cases progressing through various stages of FDS were also collected from three states. One state's data was not able to be used for modeling due to limited reporting, but data from the two remaining states (including two state-level labs and one local-level lab) provided the basis for understanding the likelihood of cases progressing through the FDS stages.

¹ For the startup costs, the local crime lab estimated \$20,000 in labor hours for developing and validating its FDS software and a new laptop (purchased for speedier processing). They also reported \$50 in labor costs for each of fifty staff members to attend training as well as approximately \$500 in labor and supplies costs for developing the training materials (for a total training cost of \$3,000). Ongoing costs for each individual FDS request included an estimated 22 labor hours spent should the FDS request go through all stages of the process—i.e., the request process, software search, lineage testing, review of non-DNA information/records, and sharing investigative lead information with police. In addition, there were ongoing supply costs for lineage testing, including approximately \$500 for reagent kits for testing an individual evidence sample against 10 CODIS reference samples and bulk costs of about \$22,000 for a batch of kits for testing 50 evidence samples against 500 CODIS reference samples.

METHODOLOGY (continued)

The cost model framework builds on the Markov Decision Process (MDP) that requires (at minimum) two key sets of inputs from a user—(i) a set of stages that a FDS request can pass through and estimates of the transition probabilities and (ii) monetized costs (and, where available, benefits) associated with each stage. While Markov Decision Processes can vary in complexity, the general framework makes it easy to compute expected costs associated with all stages of a process while properly discounting costs associated with states that are less likely to occur. More elaborate variants of the MDP can include costs and benefits of certain actions associated with each of these states. The current application uses a simplified version of the MDP.

Because not every FDS request may result in an FDS search, and not every search yields a match, a case has an increasing likelihood of dropping out as it proceeds through the system. The MDP transition probabilities (or likelihood estimates) provide a simple means of discounting costs associated with those later stages. Below illustrates the calculations involved in the model.

Let the likelihood of an FDS case reaching stage j be denoted by p_j and let the variable per unit cost associated with each stage be denoted by c_j . Then, the total probability for an FDS case request resulting in the release of the identity of a potential familial relationship to law enforcement authorities can be computed as:

$$P = p_A + p_B + p_C + p_D + p_E$$

The expected total cost (including costs incurred in all future stages) associated with any FDS search can be computed as:

$$C = c_A * p_A + c_B * p_B + c_C * p_C + c_D * p_D + c_E * p_E$$

The expected cost needed to generate a lead (suspect ID or potential familial relationship) can be computed as:

$$C^* = C/P$$