
UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970

Escobedo, Israel, Validation of the Laboratory Information Systems Applications (LISA): Kinship Analysis Module. Master of Science (Biomedical Sciences), December 2008, 155 pp., 10 tables, 21 figures, references, 34 titles.

The Laboratory Information Systems Applications (LISA) is a LIMS system designed for the management and analysis of genetic data from large scale investigations of human identification. The validation study evaluated the pedigree construction and analytical tools of the Kinship Analysis module. The genetic data from old paternity cases was used to construct pedigrees under several scenarios that simulate situations involving a missing child/parent. Each pedigree was analyzed to obtain a KI that measures the strength of the observed genetic evidence for an association made between a missing/deceased individual and a family reference pedigree to make identification. A common distribution of the number of observations per range of the ln KI was observed in all scenarios. A concordance and reproducibility study was conducted for eight families using a validated kinship analytical program. LISA has the potential of becoming an efficient tool for human identification despite the limitations of the software.

VALIDATION OF THE LABORATORY INFORMATION SYSTEMS APPLICATIONS (LISA): KINSHIP ANALYSIS MODULE.

INTERNSHIP PRACTICUM REPORT

Presented to the Graduate Council of the

Graduate School of Biomedical Sciences

University of North Texas
Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Israel Escobedo, B.S.

Fort Worth, Texas

November 2008

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AF – Alleged Father

AFDIL – Armed Forces DNA Identification Laboratory

AFR – African American

AMX – Admixture

C – Child (Sibling or Offspring)

Cal DOJ – California Department of Justice (Bureau of Forensic Services)

CAU – Caucasian

CODIS – Combined DNA Index System

DM – Double Match Score

DNA – Deoxyribonucleic Acid

FKAs – Fortuitous Kinship Associations

FTI© - Future Technologies Incorporated

G - Genotype

HIS – Hispanic

IBD – Identical By Descent

IBS – Identical By State

KADAP – Kinship Analysis and Data Analysis Panel

KI – Kinship Index

LIMS – Laboratory Information Management System

LISA – Laboratory Information Systems Applications

ln KI – Natural Logarithm of the Kinship Index

M –Mother or Mutation Score

mtDNA – Mitochondrial DNA

P – Parent/Spouse or Probability

PI – Paternity Index

QA/QC – Quality Assurance and Quality Control

R – Rare Allele Score

RFLPs – Restriction Fragment Length Polymorphisms

SM – Single Match Score

SNPs – Single Nucleotide Polymorphisms

SOPs – Standard Operational Procedures

STRs – Short Tandem Repeats

UHRs – Unidentified Human Remains

VNTRs – Variable Number Tandem Repeats

WTC – World Trade Center

Y-STRs – Y Chromosome Short Tandem Repeats

CHAPTER I

INTRODUCTION

Kinship analysis is an important tool in the identification of human remains found in mass disasters (natural, accidental, or terrorist attack), mass grave investigations, or missing persons' cases. The objective of these situations is to identify the unknown human remains in order for the medical examiner to sign a death certificate and quickly return the remains to the grieving families of the deceased individual for burial [23, 26]. The identification of human remains is first carried out using traditional methods of human identification from the fields of forensic anthropology, radiology, odontology, and fingerprint analysis. Using such methods of human identification cannot guarantee successful identification of human remains. The state of degradation, decomposition, and fragmentation of human remains can interfere with identification using traditional methods. However, DNA based methods of identification have shown to be more reliable than traditional methods when identifying human remains because it requires small amounts of biological sample from human remains in any condition and DNA is unique to individuals except with monozygotic twins. The objective of DNA based methods is to obtain a DNA profile consisting of genetic markers like short tandem repeats (STRs) from human remains of an unknown individual and compare it with a DNA profile from a reference sample representing the victim or missing person to make the match. Once a match is made between DNA profiles, statistical weight is added to the match to demonstrate the uniqueness or relevance of such match [20, 23].

Making a match between DNA profiles of a questioned sample obtained from human remains and a reference sample is an ideal situation which is not the case in most investigations involving mass disasters, mass graves, and missing person cases. The root of the problem is the quality of the DNA profile recovered from both questioned and reference samples. The quality as well as the recovery of DNA profiles of questioned samples depends on the condition of the human remains. In mass disasters and mass grave cases, human bodies are exposed to several elements (fire, water, wind, chemical insults, or physical forces) and the amount of time exposure to such elements can affect the rate of decomposition and degradation of human remains. The effects of such elements for a period of time on human remains can interfere with the recovery of sufficient amounts of DNA for analysis and leads to the formation of partial DNA profiles caused by degraded or low copy amounts of DNA found in questioned samples. Another effect is the formation of mixture profiles which occurs in situations of comingling between body parts of different individuals in a mass grave or mass disaster. Mixture profiles obtained from questioned samples are usually avoided or reanalyzed to get a better profile. New technologies can help improve the recovery and quality of DNA profiles obtained from questioned samples like mini-STR primers that can help improve the quality of STR profiles. STR testing is the most common form of DNA testing for questioned samples but the difficulties of obtaining complete STR profiles are problematic and require the use of other informative genetic markers that are found in large quantities in degraded biological samples. Such genetic markers consist of the hypervariable regions of mtDNA which provides information of maternal lineage of

individuals, Y-STRs which provides information of paternal lineage of individuals, and single nucleotide polymorphisms (SNPs). However, mtDNA and Y-STR analysis cannot provide individualizing information while SNP testing requires assaying a large number of SNP sites to make it statistically discriminating for identification [9, 23].

As for reference samples, the question is not whether a good quality DNA profile can be obtained from such a sample but whether the reference sample truly represents the missing individual or victim. The majority of cases involving the identification of unknown human remains through DNA analysis require the use of reference samples. One of the best and most reliable categories of reference samples are clinically preserved archival biological tissues like paraffin embedded tissues or bloodspot cards because traceable medical or legal records exist to demonstrate that the preserved tissues in fact comes from the victim or missing individual. The second category of reference samples are surrogate samples or biological samples found on personal items presumed to come from the victim or missing individual. Personal items can consist of hairbrushes, toothbrushes, envelopes, stamps, toys, pacifiers, razors, bedding, and other items that may have DNA from the individual in question. Many personal items are obtained from the victim's or missing individual house or from the crime or mass fatality scene. The problem with surrogate samples is its reliability because it can generate a DNA profile that is completely different from the actual victim, mixture profiles, or no DNA can be recovered from such items. For example, in the identification of victims of Swiss Air Flight 111 disaster 10% of the 250 personal items used to obtain a reference DNA profile produced mixture profiles or profiles different from the DNA profiles of the remains of

the victims. Many of the problems with surrogate samples are attributed to errors in the sampling process especially when family members of the victim with no prior experience in evidence gathering are involved increasing the potential of contamination. If the reference sample is found in the scene of the crime or a disaster, then the problem is degradation by environmental insult. Confidence for a reference sample can increase if the DNA profile is obtained and observed in multiple personal items [8, 9, 20].

However, the biggest problem regarding reference DNA samples for the identification of human remains is the existence of such samples. References samples may not exist for any reason or the recovery of DNA from personal items was unsuccessful. The nonexistence or the lack of reference samples can also occur in situations like disasters or attacks in which entire families and their belongings are destroyed impeding the process of recovering reference samples from their homes or relatives. The best example of this situation is the East Asian tsunami of 2004 in which entire families, communities, and homes were destroyed making the recovery of reference samples or contacting relatives difficult in the identification process of the victims [4, 11, 23].

The lack of references samples is a common problem in identifying human remains; hence a different approach is used for identification call kinship analysis which requires the use of DNA profiles from relatives of the missing person or victim. Kinship analysis is an indirect statistical approach used in identifying human remains when there is no reference sample available or the reliability of the reference sample is in doubt. The approach is to find an association between DNA profiles of the unknown human remains

and the relatives of the victim or missing individual in question. The idea is that if the DNA profile of the human remains really comes from the victim or the missing individual from a particular family, then it will share alleles common with that family or are identical by descent (IBD). If an association can be made between the DNA profile of the human remains and those of the relatives of the family in question then statistical methods can be used to obtain a kinship index (KI) which is a value that provides support for the association [24, 26]. Greater the KI value, the better the association between the human remains and the family. The number and type of relatives used in kinship analysis can influence the value of the KI. For example, parents of a missing individual can provide better information for the identification of human remains compared to a sibling unless there is large number of siblings that can be tested [8, 23]. Although having first degree relatives like parents, siblings, and offspring for kinship analysis are the best but circumstances may direct kinship analysis to look at second degree relatives like grandparents, half siblings, uncles/aunts, cousins, or nephews/nieces. The use of second degree or even third degree relatives may be used in kinship analysis when entire families are wiped out like in the Swiss Air Flight 111 Disaster or the East Asian tsunami of 2004 [4, 20]. The availability of DNA reference samples from first and second degree relatives can be problematic if the victim's living biological relatives are few or are unable/choose not to participate in the identification process [23]. The problem worsens if the reference sample from a relative has a differing biological or nonexistent relation to the missing individual/victim which can occur with inaccurate, conflicting, or undisclosed

information on the part of the family. It is important on the behalf of the investigators to acquire accurate information to avoid inconsistencies in kinship analysis [23].

Obtaining a KI for an association between a family and questioned samples requires complex mathematical calculations requiring the use of computer software. The computer software must not only calculate the KI for an association but carry out search algorithms to make a match between questioned samples and known family reference pedigrees using a combination of genetic data like STRs, mtDNA, Y-STRs, and SNPs. However, a problem arises when using computer programs for kinship analysis in which KIs can be obtained for associations between family pedigrees and the genetic profiles of individuals unrelated to the family; this occurrence is known as FKAs (fortuitous kinship associations). FKAs can occur when an unrelated individual shares common alleles with those of an unrelated family by chance or identical by state (IBS). In the Swiss Air Flight 111 accident investigators expected 1 in every 2000 pairwise comparisons produce a KI for parent/offspring relationship that is nonexistent. Investigators from the 9/11 WTC attack expected to observed 13,000 FKAs after 26 million searches between a questioned sample and (single/double) next of kin reference sample. The KI obtained from an FKA can put in doubt the association made between the human remains and the family reference pedigree. The FKAs appear in situations when the association is made with a family pedigree that consisted of one relative or a small number of distant relatives. The FKA problem demonstrating the need of obtaining DNA profiles from many relatives as possible. Not only would the statistical weight placed on the association between a set of remains and a reference family pedigree be in doubt but the possibility of

misidentification can be legally and financially devastating to the laboratory, grieving family, law enforcement agencies, and the public. To ensure the reliability of the KI when an association is made between a set human remains and a family reference pedigree there must be a threshold value or a range of KIs that can successfully discriminate between false and true associations [19, 20].

CHAPTER II

BACKGROUND

Kinship analysis has its basis in paternity analysis in which genetic data is used to determine the father of a child. In most paternity cases, STR data is obtained from the child, the mother, and the alleged father to be used for determining the alleles of the child's STR profile that are shared between the mother and the alleged father. According to Mendelian inheritance, the child's STR profile must share half of the alleles with the mother and the other half from the biological father which may be the alleged father. If a match or an association is observed between the alleles of the child and the alleged father then statistical weight can be added to the observation. The most common statistic used in paternity analysis is the paternity index (PI) which is a likelihood ratio of two probabilities (PI = X/Y). Each probability (X and Y) represents hypothesis about the relationship between the child and the alleged father based on the observed genetic profiles or each individual. The numerator (X) is the probability that the alleged father is the biological father based on the observed genetic profiles and the denominator (Y) is the probability that a random individual is the biological father based on the observed genetic profiles. For example, the PI for an STR locus in a situation in which the mother has a CD genotype, the child has a BC genotype, and the alleged father has an AB genotype; the ratio would be $1/(2P_B)$. The X is equal to ¼ because it is the probability that the offspring would have a BC genotype if the mother and father have genotypes CD

and AB respectively. Y is equal ($1/2 \times P_B$) is the product of the maternal contribution which is half and the allele frequency of the B allele in the population that represents the paternal contribution of the B allele to the offspring. The X and Y equations are divided resulting with the equation is $1/(2P_B)$. Similar equations for the PI can be obtained the same way depending on the homozygous or heterozygous genotypes of the child, mother, and alleged father for a particular locus [8, 16].

The PI of many tested loci can be multiplied according to the product rule to combine them into a cumulative PI that describes the strength of the genetic evidence for the paternity. The application of the product rule can also be used to combine the PI and the likelihood ratios of other independent genetic systems like mtDNA or Y-STRs to increase the weight of the genetic evidence. A high PI value around the thousands provides strong support for paternity but a low PI value provides weak or no support for paternity. Most non-paternities will have low PI values near, at, or below one which occurs when Y is greater than X in the PI ratio. The combine PI can be increased if the number of tested loci is increased but if non-paternity exists then testing more loci will have little effect on the overall PI. The calculations for the PI can be modified to take in account linked loci, mutation, and population substructure. Modifications for PI calculations for linked loci are only used when using genetic systems with observed linkage disequilibrium like the HLA haplotype system. PI calculations modified to account for mutations are only used when inconsistencies are found between the STR profiles of the child and the alleged father to prevent false exclusions. To account for mutation in the PI calculation, the mutation rate (μ) of the STR loci where the

inconsistency is found is incorporated into the calculation which leads to a reduction in the PI [8, 16].

Modifications that account for population substructure in PI calculations are usually done in situations in which the mother and father share a common coancestry because they live in a small isolated community. A coancestry coefficient referred as a theta value (θ) which is specific to a particular population is incorporated into the calculations for the PI to account for population subdivision. The θ value is related to Wright's F_{ST} value and is usually known as the probability of two homologous alleles that are IBD for a locus. The θ value is known to reduce the PI but the effect on the PI of a paternity case from a large population is quite small. The reason is that the PIs of paternity cases from large populations are very high that the reducing effects of the θ value are negligible. Most laboratories do not use θ value for determining PI for routine paternity testing and have reserved it for special cases. In some rare situations, the incorporation of the θ value can significantly reduce the PI for a paternity case. For example, in one case from Hong Kong in which the PI provided support for a true paternity until θ values of .01, .02, and .03 were used to recalculate the PI. The incorporation of θ values caused the PI to drop significantly providing support for non-paternity than a true paternity. According to Fung et al, the Hong Kong case outlines the need to choose the correct θ value depending on the situation of the case especially for immigration cases and parentless cases [1, 15].

The PI can be reduced if the mother's genotype or genetic profile is unknown because the calculations for determining motherless paternity change. For example, if a

child's genotype for a particular locus is AB and the alleged father's genotype is AC for the same locus, the PI calculation would be $1/(4P_A)$ as supposed to $1/[2(P_A+P_B)]$ if the mother's genotype for the locus is AB. The X is the product of the probability of transmission of the paternal allele which is half and the allelic frequency of B in the population to represent the unknown mother. Y would be the child's genotype frequency in the population which is $2P_AP_B$. The resulting PI equation when X and Y are divided is $1/(4P_A)$. Depending on the homozygous or heterozygous nature of the genotypes of the child and alleged father, the PI calculations for motherless paternity can be different or identical to PI calculations for a regular paternity case [8].

Obtaining a PI from a motherless paternity case must be taken with caution because of the possibility of a false paternity is high without knowing the mother's genotype or genetic profile. In one study by Schwark et al, the PI was obtained from 93 child/biological father pairs and 125 child/uncle (brothers of biological fathers) pairs. The PI values for child/biological father pairs were higher compared to PI values from child/uncle pairs. Despite the low PI values from child/uncle pairs, only 5 child/uncle pairs with non-excluding loci had PI values high enough to consider the uncle to be the father of the child based on international standards that require a minimum PI =1000 for the alleged father to be included as the father of the child. Similar results were obtained with child/uncle pairs with one or two exclusionary STR loci out of the 15 loci being tested. Same thing occurred when exclusionary STR loci were ignored in the child/uncle pairs with one exclusionary STR loci based on international standards in which a minimum of 12 STR loci are needed to establish paternity [30]. In another German study

by Poetsch et al, STR data for 13-15 loci from 336 children and 348 unrelated men to analyze motherless paternity cases. Using computer software to compare the STR profiles of children with those of unrelated men to make an association or match resulted with 116,004 child/man pairs. The results showed that 322 children had at least one unrelated man that could not be excluded as the father of the child and 160 children had more than one unrelated man that could not be excluded as the father. In both groups of children, the unrelated men that were included as the father of the child had 3 to 0 exclusionary STR loci, and 1666 pairs out of 116,004 child/ man pairs had 3 exclusionary STR loci. Every child/man pairings achieved a probability of paternity high enough to include the unrelated man as the father of the child based on German standards. The probability of paternity is a percent that represents the posterior probability of paternity after considering the genetic evidence calculated from the PI and the established prior probability [8, 24]. Both German studies show that false inclusions of alleged fathers in motherless paternity cases can occur and recommend the use of testing more STR loci to establish the paternity of the child [24, 30].

Obtaining a false inclusion or exclusion for paternity is a significant occurrence but is hard to detect even when several factors have been accounted for. The chance of obtaining false inclusion or exclusion is quite high when dealing with cases involving the identification of unknown human remains. In one study by Birus et al, pairwise searches were performed between STR profiles from exhumed unidentified human remains from a military conflict that took place in Croatia in the 1990s and a database of potential relatives of missing individuals from the war. The PI was calculated for every match

between profiles from human remains and the database. The results of that search yielded several false matches between two unrelated individuals with a high PI (20 cases with 14 loci and 4 with 15 loci) after thorough investigation of each match. The investigators of the study noticed that certain genotypes contain common alleles for a locus that can match several genotypes in the family reference database. For example, certain genotypes from unknown human remains were able to match 100 genotypes in the database for certain locus suggesting the possible presence of inbreeding. A population study of the family reference database was carried out using the GDA software. The heterozygosity and inbreeding coefficient of each STR locus in the population which are measures of inbreeding were calculated by the software. High heterozygosity and low inbreeding coefficients were observed for each locus in the database suggesting that inbreeding may not be present in the population. A second study was performed using a mini-haplotype approach by producing all the possible 3 loci combinations of STRs and calculated the expected frequency of 3 loci combinations in the population to be compared with the observed frequency to find any deviations that may be the result of inbreeding. The results showed that the average occurrence of certain mini-haplotypes was 4.8% in the population with the largest with 27.7%. Several mini-haplotypes had significant deviations from the expected frequency suggesting the presence of inbreeding. The investigators suggest that the family reference database comprises of inbred individuals from small isolated communities in Croatia and that the mini-haplotype approach can help identify inbreeding in small sub-populations [7].

Determining the parentage of an individual using paternity testing is very useful in identifying human remains but what if the investigation requires determining the type of familial relationship between two individuals like full/half siblings, cousins, nephews/nieces, grandchildren or other. The approach of determining relatedness between two individuals takes advantage of allele sharing between relatives. Certain relatives can share 0, 1, or 2 alleles that are IBD in common with another relative. For example, a parent and his/her child must share 1 allele in common that is IBD, while full brothers can share 0, 1, or 2 alleles that are IBD. The probability of sharing an amount of alleles (0, 1, or 2) that are IBD based on the relationship between two individuals can be obtained by looking at Mendelian inheritance. For example, the probability that full siblings sharing 2 alleles IBD can be obtained by multiplying the probability of receiving an allele from both parents (½ for each parent) is ¼ while for sharing 1 allele IBD is always going to be ½ and for 0 alleles shared is ¼. The probabilities of sharing 0, 1, or 2 alleles IBD depending on the type of familial relationship between two individuals are referred to as IBD coefficients (Z_0, Z_1, and Z_2) [8].

To determine the familial relatedness of two individuals, a conditional probability must be calculated. The conditional probability states that the probability of a genotype of an individual given the genotype of a relative P [G_1- genotype of an individual| G_2- genotype of a relative]. The probability can be broken into three parts representing the number alleles shared by IBD; P [G_1| G_2] = (P [G_1| G_2, Z_2] x P [Z_2]) + (P [G_1| G_2, Z_1] x P [Z_1]) + (P [G_1| G_2, Z_0] x P [Z_0]). The equation demonstrates that the sum of the products of the conditional probability of an event pertaining to allelic sharing between two

individuals with a corresponding IBD coefficients for a particular type of relationship being tested. To obtain the conditional probability for each of the three parts of the equation, the ITO system developed by Li & Sacks can be used. The ITO system uses three transitional matrices that represents the number alleles shared that are IBD, hence I is for 2 alleles shared by IBD, and T is for 1 allele shared by IBD, and O for 0 alleles shared by IBD. Each matrix contains probabilities for sharing alleles (0, 1, or 2) IBD specified by its own matrix for every combination of genotypes for two individuals in a bi-allelic system. For example, if two individuals have genotypes Aa and AA and share one allele IBD, using the T matrix the probability would be .5 p while under the I and O matrices it would be zero and p^2 respectively. Using the ITO system, the appropriate conditional probabilities can be plugged into each part of the overall equation. Once all the conditional probabilities and the appropriate IBD coefficients are setup according to the familial relationship being tested then plug in the allelic frequencies to obtain a likelihood of that relationship for a particular locus. Divide the likelihood for a particular familial relationship being tested by the likelihood that two individuals are unrelated to obtain a likelihood ratio. The likelihood ratio approach can be used to test one familial relationship against another familial relationship [8, 21].

Like the PI, equations for likelihood ratios that pertain to the familial relatedness between two individuals can be modified to account for factors like population substructure, and the resulting likelihood ratios can be interpreted in the same way (higher the ratio, greater the support for the existence of the familial relationship) [8]. Taking a likelihood ratio approach in determining familial relatedness between two

individuals can be difficult with the possibility of obtaining false matches, inclusions, or exclusions. In a study by Tzeng et al, likelihood ratios were calculated for 126 true sibling pairs and 126 random pairs in order to distinguish between sibling and unrelated individuals using 15 STR loci system. The likelihood ratios for sibling relatedness vs being unrelated individuals for random pairs were less than 1 and for 107 true sibling pairs the likelihood ratios were above 100. In three random pairs, the likelihood ratios were high while five true sibling pairs with likelihood ratios less than 3. MtDNA testing was used to resolve these possible false exclusions and inclusions of both categories of individual pairs. The result of the mtDNA testing showed that the inclusion of sib-ship in the three random pairs were false and the five pairs of exclusion of sib-ship were proven false as well as demonstrating that mtDNA testing can be used to distinguish false or true relationships [28].

Another problem with relationship testing is the inability to distinguish the most likely familial relationship that two individuals might have. For example, in the study by Presciuttini et al using the Profiler Plus kit demonstrated that similar likelihood ratios can be achieved in relationship testing depending on the number alleles shared by two individuals based on simulated data. The likelihood ratios were calculated under the scenarios that the simulated pair is a parent/child or full sibling pair rather than unrelated individuals. Similar likelihood ratios were obtained for both scenarios with 12 alleles shared with likelihood ratios for parent/child vs unrelated was 405.9 and for full siblings vs unrelated was 496. The objective of the study was to find a way to distinguish the type of relationships between individuals by calculating likelihood ratios for simulated

pairs and create distribution of likelihood ratios based on the number of alleles shared. The study demonstrated with 9 shared alleles all parent child pairs and 92% of full sibling pairs are identified correctly. However, at 10 and 11 alleles shared the number of identifications for parent/child pairs and full siblings depreciates [25]. In a study by Wenk and Chiafari, likelihood ratios that test for a full sibling relationship vs half sibling relationship when using three VNTR loci to distinguish full and half sibling relationships. The study showed that 18 out of 25 true full sibling pairs had higher likelihood ratios compared 23 out of 25 true half sibling pairs. However, seven true sibling pairs had low likelihood ratios comparable to those of true half sibling pairs and the same thing can be seen with three true half sibling pairs with high likelihood ratios comparable to true full siblings. The investigators determined the number of alleles shared for each pair of true full and half sibling pairs. The results showed that full siblings sharing 2 alleles in two or more VNTR loci had the highest likelihood ratios while full siblings with 1 or 0 alleles shared in all of the VNTR loci tested had the lowest likelihood ratios comparable to ratios compared half siblings. As for true half siblings, the highest likelihood ratios had 2 alleles shared in one VNTR locus while the lowest likelihood ratios that provided support against full sib-ship shared 1 or 0 alleles in all VNTR loci tested. The investigators mention that by looking at three VNTR loci is not enough to distinguish full and half sibling pairs [29]. Having two alleles between two individuals is indicative that their full siblings but observing half siblings sharing two alleles is a rare event because half siblings share 0 to 1 alleles because they only have one parent in common [28, 29].

Determining the relatedness between two individuals can be useful in identifying unknown human remains but there is a chance the resulting likelihood ratio provides support for a familial relationship that does not exist. To avoid the problem, most investigators try to find the relatedness between the human remains of unknown individual and a reference family using a method called kinship analysis. The statistical framework is the same using a likelihood ratio approach that divides the joint probabilities of two competing hypothesis that results with likelihood ratio that measures the strength of the observed genetic evidence called the kinship index (KI). The numerator represents the probability that the genotype or genetic profile of the tested individual or unknown human remains is related to a particular reference family. The denominator represents the probability that the genotype or genetic profile of the tested individual or unknown human remains is unrelated to a particular reference family. The equations for the KI are dependent on the circumstances of the case and the number of individuals being tested. Obtaining the equations for the KI is a rigorous process because their different for each case and for each locus being tested [8, 26].

For example, in a situation in which the remains of an unknown individual (G_B) is found and the evidence indicate that remains are related to a particular family consisting of one parent (G_P) and one offspring (G_C) of a missing individual. In this case, the joint probability of $P(G_C, G_B, G_P) = P(G_C| G_B, G_P) \times P(G_B| G_P) \times P(G_P)$ is the same for numerator and denominator but certain probabilities and terms must be eliminated because their two independent hypothesis. The probability of observing the genotype of the parent of the missing individual $P(G_P)$ is eliminated in both hypothesis because the

genotype of the parent is not dependent of whether the unknown individual is related to the reference family or not. The probability of observing the genotype of the offspring of the missing individual given the genotypes of the unknown individual and the parent of the missing individual $P(G_C|\ G_B, G_P)$ is modified into $P(G_C|\ G_B)$ in the hypothesis of the numerator because under this assumption the genotype of the offspring does not depend on the genotype of the grandparent which is the parent of the missing individual. In the denominator, the $P(G_C|\ G_B, G_P)$ and $P(G_B|\ G_P)$ are modified into $P(G_C|\ G_B)$ and $P(G_B)$ because under this hypothesis the parent of the missing individual is unrelated to the unknown individual. The resulting likelihood ratio is $[P(G_C|\ G_B) \times P(G_B|\ G_P)] / [P(G_C|\ G_B) \times P(G_B)]$ [26].

Assuming the genotypes for a particular locus for G_B is bc, G_P is ab, and G_C is cd; the next step would be to obtain the equations for the likelihood ratio for the case. For the numerator, $P(G_C|\ G_B)$ is equal to population allelic frequency of d divided in half ($P_d/2$) and $P(G_B|\ G_P)$ is equal to the populations allelic frequency of c divided in half ($P_c/2$). As for the denominator, the $P(G_B)$ is the population genotype frequency of the unknown individual ($2P_bP_c$) but with $P(G_C|\ G_B)$ is equal to the product of the population allelic frequencies of c and d (P_cP_d). Obtaining the equation for $P(G_C|\ G_B)$ is derived from the sum of the products of the $P(G_C|\ G_D)$ and $P(G_D|\ G_P)$ for all the possible genotypes for the missing individual (G_D) of the family. The number of possible genotypes for the missing individual is determined based on the genotypes of the parent and offspring because they share one allele IBD with the missing individual whose genotype is unknown. In the end, the resulting KI equation is $[(P_d/2)\ (P_c/2) / (2P_bP_b)(P_cP_d)]$. The example demonstrates the

complexity of obtaining the equations for the KI for a simple scenario involving the association between the remains of unknown individual and a small family with a missing individual. The KI equations become more complex when the number of tested relatives increases and the type of relationships the relatives have within the family pedigree which is a visual representation of the tested family [8].

The general concept of obtaining the KI equations is commonly used among investigators but there are several other methods that can achieve the same objective. One such method was developed by Dr. Charles Brenner in which it accomplishes the same goal of obtaining the KI equations that test the relatedness between a tested individual and a family pedigree. The difference between Brenner's method and the general method of obtaining the KI equations is that Brenner's method accounts for untyped or unknown common ancestors relevant to the family pedigree. For example, in determining the relatedness between two potential siblings the calculations for KI equations account for the untyped parents of the siblings in the pedigree. The method first determines the genotypes of each individual being tested and determining all of the possible genotypes for untyped individuals relevant to the family pedigree. Determine all of the possible combinations of genotypes if the untyped individuals are the parents. Assemble all the genotypes in a graph where each person has two columns one for genotypes and the other for probabilities. For untyped individuals, determine the genotype probabilities and for typed individuals use conditional probabilities which are the IBD coefficients. Next determine the equation for numerator of the likelihood ratio which can be done by adding up the product of the probabilities in each row. To obtain the equation for the

denominator can be done in the same manner as the numerator but this time the hypothesis dictates the tested individual is unrelated to the family pedigree meaning that the genotype probabilities will be used instead of IBD coefficients. Once the equations are obtained for the KI, division is used to further simplify the equation [3, 5]. All methods that calculate the KI used in determining the relatedness of a tested individual to a particular family are quite complex and requires the need of computer algorithms and software [8].

The equations for the KI can be modified to account for population substructure and mutation rate for special circumstances. However, incorporating the correction for population substructure in the KI calculation can bias the likelihood ratio in favor of the hypothesis in which the tested individual is unrelated to the family pedigree. Many investigators prefer not to use the correction for population substructure because the objective is to identify the human remains. The KI, like any other likelihood ratio it provides the strength of the genetic evidence for a particular proposition. A high KI provides greater support for the hypothesis in which the tested individual is related to a particular family pedigree. The KI from STR analysis can be combined with the likelihood ratios obtained from mtDNA, Y-STR, and SNP evidence by invoking the product rule to increase the support for a particular hypothesis. The resulting cumulative KI can be worded into a statement that describes the strength of the genetic evidence for particular hypothesis as opposed to the other competing hypothesis in conjunction with any available non-DNA evidence [8, 26].

The KI is influenced by the type and number of relatives being tested in the analysis as well as the relationship that the tested individual might have with a particular family pedigree. The genetic information from first degree relatives like parents, offspring, and siblings are considered the best for kinship analysis. First degree relatives can be ranked based on the amount of genetic information they can contribute to the analysis with parents at the top and half siblings at the bottom (parents > offspring > full siblings > half siblings). Using the genetic information from parents it can provide strong evidence for any inference of relationship between the tested individual and a particular pedigree because both parents must share one allele in common with their offspring. The KI would be considerably high if the genetic information from one or both parents is used in kinship analysis if the relationship between the parents and the tested individual is true. The genetic information from the offspring is also informative especially for the identification of missing or deceased parents because the offspring share one allele in common from each parent. If the missing or deceased parent has multiple offspring, it is best to test all of them because it can provide information of the allele contribution of the parent and increase the KI value for the proposed relationship if true. The inference can further be aided if the genetic information of one parent is available for kinship analysis. The use of full siblings in kinship analysis is useful but the least informative because in some cases very low KI values are observed regardless of the validity of the relationship. Low KI values are possible for full siblings because they can share zero alleles in common providing support for the alternate hypothesis. The problem can be averted if the missing or deceased individual has multiple siblings because by testing multiple

siblings it becomes easier to determine the four allele contribution of the parents. If the genetic information of the parents of the missing individual or deceased is available, it is statistically better to use the genetic information from the parents than that of siblings. The problem worsens with half siblings, the least informative of the first degree relatives because half siblings share one parent indicating they share one half of the genetic information with other siblings. The KI values are low in situations when kinship analysis of half siblings is used providing support for the alternate hypothesis. In most cases investigators avoid using half siblings for kinship analysis [8].

The identification of human remains using kinship analysis of first degree relatives cannot always be done in certain situations because reference DNA samples of first degree relatives are unavailable. In such situations, investigators are forced to obtain DNA samples from second degree relatives like cousins, grandparents, and uncles/aunts. However, the use of second degree relatives in kinship analysis is not statistically robust and informative as first degree relatives. To get around this problem it is recommended to test multiple second degree relatives or combinations of multiple first/second degree relatives in order to increase the amount information that is needed for kinship analysis to make it statistically significant. One factor that can influence the outcome of any kinship analysis is the rarity of alleles shared between relatives in a family because the rarer an allele is the more significant the associations between relatives that share the rare allele. However, the presence of null alleles in the genetic profile of the deceased or missing individual can be problematic for kinship analysis which can reduce the KI for an association between a test individual and a family pedigree. To resolve the presence null

alleles in genetic profiles, population studies that compare the allele designation for STR loci using new and currently used primer sets would be needed [8, 9, 23].

A limitation of kinship analysis is the possibility of obtaining a KI for associations between the deceased/tested individual and the family pedigree when such a relationship doesn't exist is usually referred to as fortuitous kinship associations (FKAs). The phenomenon occurs when the tested individual has alleles that are in common with the alleles shared in a family because there identical by state (IBS) not IBD. Usually the KIs for FKAs are usually low but there are some instances in which the value of the KI of an FKA cannot be distinguished with the KIs of true associations. In a simulated study by John Buckleton demonstrates how KIs can overlap between true and false pairwise relationships using 15 STR loci data from New Zealand Caucasian population. The study demonstrates that the distribution of the $\log_{10}$ of likelihood ratios for true associations between full siblings, half siblings, and cousins can overlap with those from simulated unrelated pairings. The confidence in a KI from any true associations can be reduced with more distantly related true associations. One way to avoid this problem is by testing more loci separating the distribution of KIs between true and false associations. Another way to resolve whether the KI might be from an FKA or not is by testing other genetic systems that demonstrate familial relationships like mtDNA and Y-STRs [8, 19, 26].

The problem associated with FKAs is more apparent when identifying human remains from mass disaster and mass grave investigations. The possibility of observing FKAs between a deceased individual and unrelated families is quite high in large scale investigations especially in pairwise testing. To resolve the FKA problem, investigators

prefer to use multiple first degree relatives for kinship analysis, test other genetic systems, and avoid pairwise testing but efforts have been made by researchers in trying to reduce or discriminate possible FKAs in the identification process of human remains in large scale investigations. Two approaches can achieve a reduction or discrimination of FKAs by either using screening methods or establishing a threshold KI value [6, 18, 19].

The first approach is to use screening methods for associations between the deceased individual and candidate reference families. In large scale investigations of decease individuals, the genetic profiles of multiple sources are condensed into one profile that represents the deceased individual. Next the genetic profile of a deceased individual is searched into a family reference database to find an association with a candidate family. Once the association is made, kinship analysis is performed to obtain a KI. The problem with searching between the unidentified genetic profile and a family reference database is that associations can be made with multiple candidate families with reasonable KI values. Screening methods have been developed for family pedigree searches to reduce or discriminate the number of FKAs [6].

One screening method is to develop a scoring system to rank all the possible associations from high to low relatedness. For example, investigators from the Swiss Air Flight 111 disaster developed a scoring system in order to find the best pairwise associations. The scoring system consists of three scoring categories (single match, double match, and rare allele scores). The single match (SM) score is the number of loci that share 1 allele in common between two genetic profiles. The double match (DM) score is the number of loci in which both alleles are shared in common between two

genetic profiles. The rare allele (R) score is the number of alleles in which their frequencies are below 1% in the Caucasian population shared in common between two genetic profiles. The scoring system was implemented during the pairwise searching phase of the investigation. The search results were ranked based on the highest scores that indicate exact matches or associations between a two genetic profiles. In the Swiss Air Flight 111 investigation the genetic profiles of deceased individuals was searched against a database composed of genetic profiles from deceased individuals or family references; referred as a type 1 search. If an association was made between the genetic profiles of a deceased individual and a candidate relative with a high score then a type 2 search is made. The potential relative is searched against the same database to verify the concordance between the two profiles based on the score report. A KI was calculated for the pairwise association once the identification was verified. The investigators were able to establish a threshold score using the SM scores to indicate which associations were true parent/offspring relationships. The threshold scores were specific for the type STR multiplex system used in the investigation like Profiler Plus had SM threshold scores of 8 and 9 while the combination of Profiler Plus/COfiler was 12 and 13. However, during the investigation there were some situations in which the genetic profile of a deceased individual that have more than two FKAs with scores at threshold values but such FKAs were resolved by reviewing type 2 search score results and by looking at the rarity of the alleles of the genetic profile of the deceased individual. The situations involving multiple FKAs with high SM scores showed that the genetic profile of the deceased individual was comprised of the most common alleles in the Caucasian population [18, 20].

A similar scoring system of the Swiss Air Flight 111 disaster was implemented with slight modifications in the identification process of human remains found in the aftermath of the 9/11 WTC attack. The scoring system consisted of the same three categories from the Swiss Air Flight 111 disaster investigation plus one more category which is the mutation score (M) to account for a locus in which the shared allele between two genetic profiles has a single repeat mutation. Instead of choosing a threshold score value, the investigators developed a consistency check to resolve or discriminate FKAs. The idea behind the consistency check is that if a true association exists between a deceased individual and a potential relative from a particular family with a high SM score then subsequently any pairwise comparison with other first degree relatives of the same family would have similar SM scores. If an FKA was detected then discrepancies would be seen in the SM scores of potential first degree relatives of the same family. The investigation also demonstrated that the ranking of potential pairwise associations based on the scoring system was similar to ranking based on calculated KIs [18, 19].

Other screening methods can be used to reduce the number of observed FKAs by using modified pedigree search algorithms. One approach called triangulation can be used to minimize the number of observed FKAs. Triangulation is a type of search in which potential associations are ranked or prioritized based on associations between the deceased individual and two members of a particular family. The method calculates a score to rank each potential association based on the product of the KIs for each pairwise association made between the deceased individual and each of the two members of the candidate family. The method is very quick and efficient method that prioritizes the best

candidate families for a deceased individual out of hundreds of potential reference families [6]. In the identification of human remains from the 9/11 WTC attack, the investigators originally used triangulation to find the best association in the form of a trio. Pairwise searches between the genetic profiles of the deceased individuals/family references, and within family references were used to identify all the possible trios under two scenarios known as the descending and the ascending scenarios. A trio under descending scenario consists of the victim associated with two potential parents and in the descending scenario the victim is associated with a potential parent and offspring. The investigators soon realized that the inferred relationship from the trio search was inconsistent with the self reported biological relationship to the victim in many of the family reference samples. The investigators abandoned triangulation for a systemic non-triangulated approach to find all the possible trios under two scenarios. A non-triangulated approach does not need two relatives with shared obligate alleles to find the third person in the trio. The idea behind the alternative method was to come up with all the possible trios ranked by their KIs and SM scores to identify FKAs and associations with discrepancies [18, 19].

The second approach is to establish threshold KI value that can reliably discriminate true and false associations. A Bayesian approach can be used to establish a threshold KI value in which an observed association between a deceased individual and family pedigree can reliably used to make identification. When using the Bayesian approach to establish the threshold, a posterior probability must be established which is the level of confidence to establish the identity of human remains through kinship

analysis making the chance of misidentification low. Many experts suggest that the posterior probability should be 99.9% for each identification made during the investigation. To obtain the desired posterior probability, a prior probability must be established in order to be multiplied with the resulting KI from kinship analysis for a particular association. The prior probability ($1/v$) is the probability that the human remains found during the course of an investigation belong to a particular person. Calculating the prior probability is based on the number of missing or unknown dead individuals (v) from the mass disaster or mass grave, and whether the incident was a closed or open system. A closed system is an incident in which the number of missing or dead individuals is already set and will not change based on the circumstances of the investigation. For example, a plane crash is a closed system because investigators already know the set number of dead based on a passenger manifest. An open system is an incident in which the number of missing or dead individuals is estimated or unknown, and will change based on the circumstances of the investigation and future evidence. An example of an open system would be the 9/11WTC attack investigation in which at the beginning of the investigation it was believed that the number of missing or dead was 5000 but over time with new evidence indicating that the number of dead was actually 2750. Knowing whether an incident is a closed or open system is important because any change in the number of missing or unknown dead individuals must change the prior probability. During the course of an investigation the prior probability will increase for every unknown dead individual that is identified causing a reduction in the number of missing or unknown individuals from the incident [4, 6, 9, 23].

Once the posterior and prior probabilities are established then the KI threshold can be established for the investigation. To establish a KI threshold, the inequality $1-(1-KI)^N \leq p$ must be solved using two pieces of information: (N) is the number of missing or unknown individuals from the incident, and (p) is the acceptable margin of error which is policy based. For example, if N= 5000 individuals and $p = 1/10^6$, then the KI threshold in which no mismatches or false associations are observed would be a KI $\geq$ 10 billion. However, the KI threshold must change like the prior probability when the number of missing or unknown dead individuals decreases or increase for any reason. For example, when an unknown person is identified the total number of unknown dead decreases causing an increase in the prior probability while the KI threshold decreases allowing true associations with low KI values acceptable for identification [9, 23].

The second limitation regarding kinship analysis is that it is computationally intensive and requires the use of computer software to carry out the calculations. Many computer programs used for kinship analysis have many specific functionalities like analytical tools for calculating likelihood ratios (PI or KI), pairwise/pedigree searching capabilities, profile consensus tools, screening tools, large scale data storage, graphic tools to design pedigrees, and other complex features. The bioinformatic software for kinship analysis are sold commercially or can be found as freeware over the internet but choosing the correct software for any investigation will depend on the circumstances of the case and the discretion of the investigators or investigating laboratory. However, any statistical software that will be used for kinship analysis for forensic purposes must be validated before use in actual casework [8, 9, 23, 32].

One of the most commonly used computer software packages for kinship analysis of large scale investigations is DNA·VIEW™ developed by Dr. Charles Brenner. The program was developed in 1988 to serve as an advanced analytical tool that can be used by both forensic and paternity labs world wide. Since the inception of DNA·VIEW™, the software has been used in some of the most important large scale human identification projects around the world like the 9/11 WTC attacks, the war mass graves found in Bosnia and Croatia, East Asian tsunami of 2004, and the Swiss Air Flight 111 disaster. DNA·VIEW™ is a multifaceted analytical program that can be used for any type forensic or paternity investigations and research. The program can use a variety of data like STRs, SNPs, Y-STRS, RFLPs, polymarkers, and X-linked markers that can be imported manually or automatically using the data files from programs like GeneMapper ID©, or Genotyper©. The functions of the program can be divided into six modules with reporting capabilities: crime case calculators, paternity case analysis, kinship, disaster identification, population analysis, and databases. The crime case calculator module is used for finding the frequency of a genetic profile and for complex mixture calculations. The paternity cases analysis module is primarily used for calculating paternity statistics like the PI. The kinship and disaster identification modules are similar because both have kinship analytical tools especially for complex cases involving inbreeding and pedigree reconstruction, searching capabilities, and applications for kinship simulations. Both modules use Brenner's method of calculating the KI which is a recursive algorithm that calculates the likelihoods of two competing hypotheses by calculating all the possible genotypes of each person in a pedigree and their corresponding probabilities. The

population module is used for calculating population statistics, estimating mutation rates of loci, and performs tests for Hardy Weinberg equilibrium, independence, and similarity between populations. The database module is for selecting, adding, deleting, combining, and managing DNA databases [2, 5].

DNA·VIEW™ is the most widely used statistical program in the world for DNA analysis with several advantageous features like it's compatibility with most operating platforms (Windows XP or Vista). However, the program has several disadvantages like its incompatibility to Mac operating systems, the fact that DNA·VIEW™ is a MS-DOS based program, hard to use, expensive (more than $7500), difficulties in printing reports, does not analyze mtDNA data, no graphical design tools for constructing pedigrees, and other problems. Due to the disadvantages related to DNA·VIEW™, many researchers and software designers have tried to design comparable user friendly programs or Excel® spreadsheets that use the kinship algorithms created by Brenner [2, 10, 27]. For example, PaternityIndex is a simple graphic interface program that allows the user to calculate the KI for any association using Brenner's method. The program with easy to use menus can import data using ASC II files, design pedigrees, calculate PI, perform multi hypothesis analysis, and imports DNA profiles onto a pedigree [10]. Another example is DOJ VIEW Version 2.0 which is an Excel® spreadsheet developed by Steve Myers from Cal DOJ that uses Brenner's method to calculate the KI for associations for families up to 10 relatives. It also has pedigree display that highlights relatives being analyzed, reporting applications, and simple database management capability [22].

Commercially available software packages contain several capabilities and features that make such programs advantageous for laboratory use. One example of these commercially available programs is GeneMarker® HID made by SoftGenetics® LLC which has several easy to use tools and capabilities for relationship testing, pedigree construction, searching, and database management. The program also acts as replacement or alternative for genotyping programs like GeneScan®, Genotyper®, and GeneMapper ID® meaning that it can accept data from many of the common Genetic Analyzer models made by ABI PRISM®. Another main feature of GeneMarker® ID is that it acts as an expert system for mixture analysis featuring deconvolution tools that can be used at the electropherogram level. The only drawback with such commercially available programs is that their expensive but many of these software packages can include warranties, technical support, service agreements, online training, free manuals, and automatic updates which are advantageous incentives for a laboratory considering purchasing such programs [27, 39].

Most programs and Excel® spreadsheets that are freely available in the internet but have limited capabilities compared to commercially made software. For example, the Excel® spreadsheet PATCAN can only do paternity statistics and likelihood ratios for pairwise relationship [27]. A second example would be FAMILIAS, a program designed by Egeland and Mostad for the sole purpose of choosing the most probable pedigree from a set of pedigrees using a Bayesian approach to calculate the likelihood ratio. The program allows the user to generate a set of possible pedigrees based on the genetic information of the tested individuals. A prior probability is established for a distribution

of possible pedigrees for a scenario(s) based on non-genetic evidence. The program calculates likelihood ratios and posterior probabilities for each pedigree to find the most probable pedigree. A multi hypothesis analytical program like FAMILIAS is designed for situations in which investigators are building family pedigrees out of scratch. Such a situation occurs in large scale investigations in which entire families are found dead and the investigators only have the genetic data of the family but lack information about the interrelationship of the individuals in the family [11,12]. Another example would be the software package developed by Wing K. Fung that consists of four separate programs each with their own special statistical features known as EasyPA, EasyPAnt, EasyIn, and EasyMISS. All four programs contain easy reporting capabilities, easy to visualize built in pedigrees with pull down menus to input data, the capability to set the prior probability for using the Bayesian approach in calculating likelihood ratios, and multi hypothesis testing. However, the software package lacks the tools for designing complex pedigrees, contains no searching capabilities, and no flexible data importing capabilities [14].

Prior to the tragic events of 9/11, most large scale investigations for human identification were carried out using multiple computer programs and Excel® spreadsheets each with unique features to compliment as a whole for kinship analysis. For example, two computer programs were used in the investigation Swiss Air Flight 111 disaster; DNA·VIEW™ and Kinship Analysis developed by Visual Basic. Kinship Analysis program was used to condense the genetic profiles from multiple sources into one profile, pairwise searching, identifies rare/mutant alleles, and generates score reports for search results. It was until the magnitude of the challenges to recover and identify

human remains from the WTC attack site that issues regarding the process and methods involving human identification came into scrutiny. The National Institute of Justice and other government agencies created KADAP (Kinship Analysis and Data Analysis Panel) which is a group of experts in the fields of forensic genetics and statistics to provide advice for the WTC identification efforts and recommendations to help forensic laboratories design a better response plan for large scale investigations. The KADAP recommendations were published in a report called "Lessons Learned From 9/11: DNA Identification in Mass Fatality Incidents". The report mentions several recommendations regarding new DNA technologies, identification strategies, project management, sample handling, statistical approaches, and other relevant issues [9, 20, 23].

The KADAP report also makes recommendations regarding information technology and biostatistical software that can be used by a lab for human identification efforts in large scale investigations which is the most important overlooked aspect by forensic investigators. KADAP recommends that forensic laboratories should adopt a Laboratory Information Management System (LIMS) that is separate and independent from the lab's regular casework LIMS system for large investigations. A LIMS system is a computer software system that stores, manages, organizes, tracks, and analyzes large quantities of data making such computer systems suitable for large scale investigations. A separate LIMS system would reduce the number of errors committed by the lab during the identification process, it would organize better the investigation, cut time within the investigation, and would not interfere with routine casework. Such a separate LIMS system for large scale investigation will contain special features that most current

laboratory LIMS systems do not have. KADAP recommends that a LIMS system for large scale investigation should have user friendly capabilities to track samples, import/export profiles in several formats, retrieve/record sample data, integrate with other software systems, search/match/combine genetic profiles from multiple sources, calculate statistical metrics, store metadata, build family pedigrees, search pedigrees, calculate likelihood ratios for associations, perform multi hypothesis testing, manage population databases, contain prioritization/screening schemes, perform technical/administrative/ (QA/QC) reviews, write reports, provide a means to communicate information between/within labs, and other important features. The implementation of a separate LIMS system allows the laboratory to organize an investigation in efficient and expedient manner by having all the bioinformatical tools in one high throughput package. Any implementation of any commercially or in-house LIMS system must be built prior to any incident/investigation, tailored to the needs of the lab, and must be validated before use in any investigation [9, 23].

One example of a LIMS system that fulfills many of the recommendations mentioned in the KADAP report is LISA (Laboratory Information Systems Applications). LISA was developed by Future Technologies Incorporated (FTI)© in collaboration with the Armed Forces DNA Identification Laboratory (AFDIL) for using large quantities of genetic data obtained from 938 specimens and 104 references for the purpose of identifying the human remains found in the 9/11 attack sites in Washington D.C. and Pennsylvania. The first publicly reported use of LISA was in 2007 with the identification of James B. McGovern a famous U.S. military pilot who was shot down in

Laos during a mission flying supplies to French troops at the end of the French Indochina War. Investigators from AFDIL obtained DNA samples recovered skeletal remains of an unknown Caucasian recovered from a possible crash site survey in 2002. AFDIL investigators used STRs, Y-STR, mtDNA, and SNP analysis for DNA samples from the skeletal remains and family references. Using LISA, the investigators managed to construct a pedigree that represents the association between the skeletal remains and McGovern's family, and then perform kinship analysis for the association. The resulting KI of 96,900 provided enough support to show the skeletal remains were related McGovern's family as suppose to the alternate hypothesis of being unrelated. Due to the success of LISA with AFDIL that many laboratories in the U.S. like Cal DOJ, Louisiana State Police Crime Laboratory, University of North Texas Center for Human Identification, and others will implement tailored LISA systems for large scale human identification projects [13,17, 33].

LISA is a multifaceted program that contains several applications that are used throughout the human identification process. The LIMS system consists of many applications like Case Management, Lab Processing, Analytical, and Lab Configuration. The Case Management is an application is used for case accessioning and tracking at the beginning of an investigation. Case information like the case/agency/lab numbers, case documents, evidential pictures, sample type, status reports, family information, medical records, and other metadata regarding the case can be entered using the application. The Lab Processing application is used to monitor the status of evidential/reference samples throughout the genetic testing process that consists of the evidential screening, DNA

extraction, quantification, amplification, and electrophoresis. The application has the capability to set up sample plates, and access worksheets/standard operational procedures (SOPs) [31].

The Analytical application of LISA is where most of the kinship analysis and searching tools are located in the system. The Analytical application consists of six modules: Administrative, Profile Management, Statistical Analysis, Searching, Kinship Analysis, and Mixture Statistics. To begin an investigation using LISA, the analyst must use the Lab Configuration application an administrative tool used to create a case set which serves as a unique storage area in the system where only the information relevant to the investigation is placed in. The information placed in the case set may consist of genetic profiles, mixture profiles, and pedigrees. Once the case set is made, the analyst can begin high throughput importing of genetic profiles into the appropriate case set in LISA using the Profile Management module. The module can manually or automatically import STR, mtDNA, Y-STR, and SNP profiles into LISA using Tab Delimited formats or XML formats from GeneMapper ID©, Genotyper©, Sequencher©, CODIS. The Profile Management module allows the analyst to export profiles in any format, add comments about the sample, modify sample names, select sample types, change identification/case numbers, delete/edit profiles, and select the type of amplification kit used on the sample. After importing profiles in the appropriate case set in LISA, the analyst can begin the data analysis portion of the investigation by selecting the Statistical Analysis, Searching, Kinship Analysis, or Mixture Analysis modules. Each module contains the capability to

generate reports with the exception of the Profile Management and Administrative modules [31].

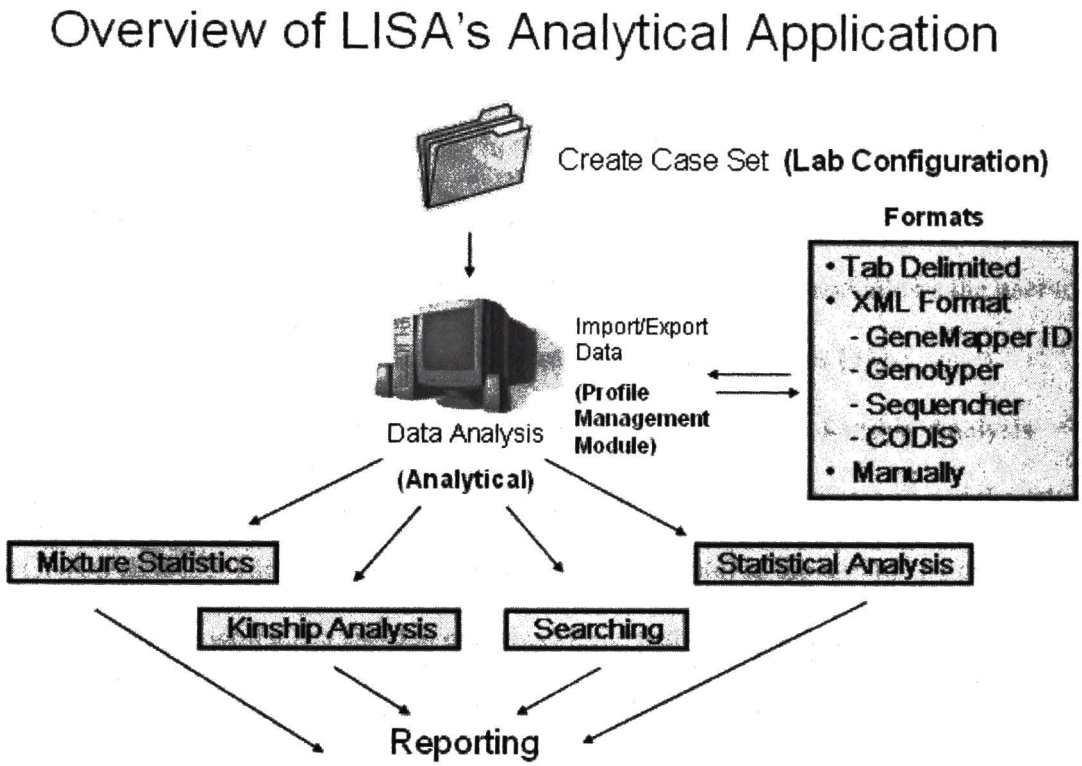


Figure 1: A complete overview of LISA's Analytical application during the human identification process.

The Administrative module is only used for database management and establishing analytical settings. The Statistical Analysis module is used for the purpose of obtaining the random match probability and the frequency of the selected STR profile plus it can calculate the paternity index, probability of paternity, and pairwise relationship likelihood ratios for selected profiles. The Mixture Statistics module allows the analyst to perform complex statistical analysis of mixture samples and calculates exclusion probabilities and likelihood ratios. The UHR vs Pedigrees allows the analyst to select the

UHR profile (a profile from unknown deceased individual) to be searched among all of the pedigrees built in the investigation file(s) using the Kinship Analysis modules. The results of the search are ranked by the highest likelihood ratio calculated for each association. The Search Profile Suite component allows the analyst to select any profile to be searched among all other profiles in the case set(s). The search can be modified too look for direct profile matches, children, parents, and first/second degree relatives. Again any match made is ranked based on the calculated likelihood ratio and allows the user to see the profiles that matched to determine rare/mutant/shared alleles [31].

The last module in the analytical application of LISA is the Kinship Analysis module which uses two integrated computer programs one called Progeny© and the other is DNA·VIEW™. The analyst can design a new pedigree(s) in the module by creating an investigation file which is a file that contains all the pedigrees built for a particular case in the case set. Once the investigation file is created, the analyst can proceed to use the tools provided by Progeny© to design complex pedigrees in a simple manner. The module allows for the analyst to select any STR, mtDNA, Y-STR, and SNP profiles in order to import them onto specific individuals in the pedigree. The individuals in the resulting pedigree will color coded based on the type of genetic markers that each individual has. To perform kinship analysis, the analyst must select any target individual in the pedigree and use the simple interface with DNA·VIEW™ to run the analysis without dealing with the archaic programming of MS-DOS. Once kinship analysis is completed, the results are presented as KIs or likelihood ratios for the four genetic marker systems and

cumulative for each of the three major population groups: African Americans, Caucasians, and Hispanics [31].

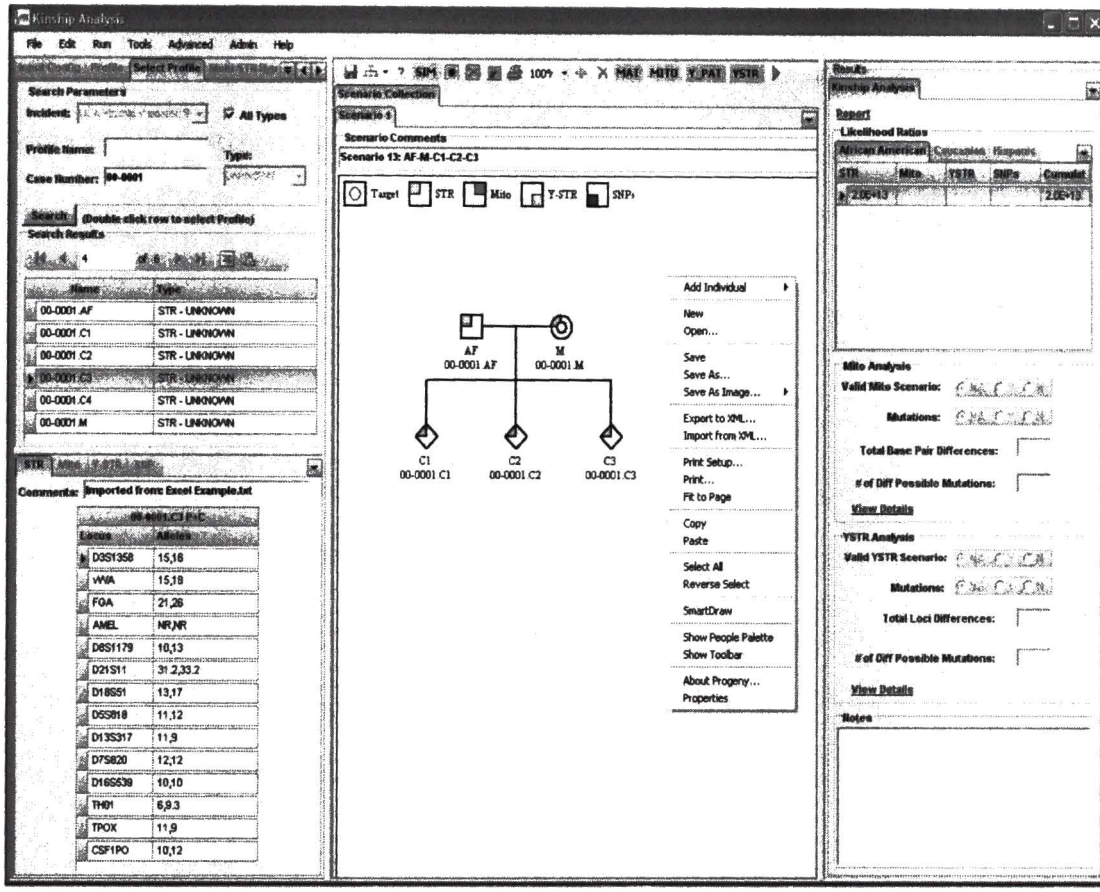


Figure 2: An example of an investigation file in LISA's Kinship Analysis Module.

Other features in the Kinship Analysis module is the capability to build/duplicate alternate pedigree, naming pedigrees, adding comments to each pedigree, perform multi hypothesis testing, setting a prior probability, smart drawing, (automatically corrects misshapen pedigree), view profiles by selecting individuals in pedigrees, blocking

profiles for individuals in the pedigree, and much more. Another feature about the Kinship Module is the detail reports it generates that contain the pedigree, the profiles of the tested individuals, calculations for the KI for each loci and cumulative for the three major population groups. A LIMS system like LISA can greatly improve a laboratory's efficiency and organization when dealing with large amounts of data commonly seen large scale investigations of human identification [31, 33].

CHAPTER III

OBJECTIVES

The University of North Texas Center for Human Identification is implementing a new LIMS system called LISA (Laboratory Information Systems Applications) designed by Future Technologies Incorporated (FTI)© for the purpose of storing, managing, and analyzing large quantities genetic data from large scale investigations [13]. A validation study of LISA's Kinship Analysis module will be conducted in compliance with quality assurance standards established in the field of forensic DNA analysis; in which any new method, software, or technology implemented by a forensic DNA laboratory must be tested and documented for future use in casework [33]. Below is a list of objectives for the validation study.

- Evaluate import and exporting capabilities of DNA profiles using the Profile Management module.

- Test the pedigree construction and kinship analytical tools for the Kinship Analysis module by building and analyzing family pedigrees under several scenarios.

- Conduct concordance and reproducibility of kinship data compared with the results from a validated analytical program.

- Analyze kinship data (mean, range, natural logarithm (ln) of the KI, and the number of observations per range of the ln KI).

- Report any problems and recommendations with LISA to FTI©.

- Write protocol for using LISA.

- Goal: Obtain a distribution of the KI or the ln KI for each scenario in order to be compared to the KI or ln KI distribution of false hits that will be obtained in a future study involving the Searching module to find a threshold value or range of values where true and false kinship associations are distinguished.

CHAPTER IV

MATERIALS AND METHODS

A validation of the Kinship Analysis module of the LISA software was conducted using old paternity cases (N =139) investigated by the University of North Texas for Human Identification. All paternity cases contain STR data for each individual of the 13 core loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX, and CSF1PO). Some paternity cases contain STR data for Amelogenin, Penta D, Penta E, D19S433, F13A01, FESFPS, F13B, LPL, and D2S1338 loci but these loci were excluded from the calculation of the KI for a particular pedigree with the exception of Penta D, and Penta E. The paternity cases consist of families with one or two parents (M-mother and AF - alleged father) with two to seven offspring (C-child). Some families have identical twins or single half siblings. The population affinity of each family was determined for 121 out of 139 families being analyzed. The remaining 18 families were categorized as admixture families because the members of such families come from different racial backgrounds. The chart (Table 1) in the following page shows the demographic data of each family that was used in the study.

Family Demographics							
Type Of Families		C1-C2	C1-C3	C1-C4	C1-C5	C1-C6	C1-C7
N		1	87	34	11	5	1
Population Groups	AFR		35	8	3	1	
	CAU		12	5	2	1	
	HIS		31	18	6	3	1
	AMX	1	9	3			
Single Parents	M		2			1	
	AF		4	1	2		
Half Siblings			18	7	3	2	1
Twins			1	1			
Penta D & E			62	15	4	2	
Missing Loci			1	2	1		

Table 1: The family demographics of the dataset containing the type and number of families with single parents (M- mother and AF- alleged father), half siblings, twins, genetic profiles Penta D and E, genetic profiles with single or double missing loci, and known population groups (AFR-African American, CAU-Caucasian, HIS-Hispanic, and AMX-Admixture).

The STR data for each family was manually entered into the LISA software for kinship analysis using the Profile Management module of the program. Once all of the STR profiles for each family were entered into LISA and reviewed for errors, pedigrees were then constructed under the four major categories of scenarios involving a missing child or parent using the Kinship Analysis module. The chart (Table 2) in the following page shows all of the possible scenarios for each of the four categories when dealing with a family with two parents and seven children.

Scenarios		
*Category 1: Missing child with one parent and one or more siblings.		
Scenario 1	Missing [C]	$P+C_1$
Scenario 2	Missing [C]	$P+C_1+C_2$
Scenario 3	Missing [C]	$P+C_1+C_2+C_3$
Scenario 4	Missing [C]	$P+C_1+C_2+C_3+C_4$
Scenario 5	Missing [C]	$P+C_1+C_2+C_3+C_4+C_5$
Scenario 6	Missing [C]	$P+C_1+C_2+C_3+C_4+C_5+C_6$
*Category 2: Missing child with two or more siblings.		
Scenario 7	Missing [C]	C_1+C_2
Scenario 8	Missing [C]	$C_1+C_2+C_3$
Scenario 9	Missing [C]	$C_1+C_2+C_3+C_4$
Scenario 10	Missing [C]	$C_1+C_2+C_3+C_4+C_5$
Scenario 11	Missing [C]	$C_1+C_2+C_3+C_4+C_5+C_6$
*Category 3: Missing parent with spouse and one or more offspring.		
Scenario 12	Missing [P_1]	P_2+C_1
Scenario 13	Missing [P_1]	$P_2+C_1+C_2$
Scenario 14	Missing [P_1]	$P_2+C_1+C_2+C_3$
Scenario 15	Missing [P_1]	$P_2+C_1+C_2+C_3+C_4$
Scenario 16	Missing [P_1]	$P_2+C_1+C_2+C_3+C_4+C_5$
Scenario 17	Missing [P_1]	$P_2+C_1+C_2+C_3+C_4+C_5+C_6$
Scenario 18	Missing [P_1]	$P_2+C_1+C_2+C_3+C_4+C_5+C_6+C_7$
*Category 4: Missing parent with two or more offspring.		
Scenario 19	Missing [P]	C_1+C_2
Scenario 20	Missing [P]	$C_1+C_2+C_3$
Scenario 21	Missing [P]	$C_1+C_2+C_3+C_4$
Scenario 22	Missing [P]	$C_1+C_2+C_3+C_4+C_5$
Scenario 23	Missing [P]	$C_1+C_2+C_3+C_4+C_5+C_6$
Scenario 24	Missing [P]	$C_1+C_2+C_3+C_4+C_5+C_6+C_7$

Table 2: A list of all the scenarios possible for a family with two parents and seven children is used for each of the four categories of scenarios. P-parent/spouse (mother or father) and C-child (sibling or offspring).

Category 1 scenarios involve pedigrees in which the genetic profile of the missing child is being compared to reference genetic profiles of a single parent and one or more siblings. Category 2 scenarios involve pedigrees in which the genetic profile of the missing child is being compared with the reference genetic profiles of two or more siblings. Category 3 scenarios involve pedigrees in which the genetic profile of a single parent is compared to the reference genetic profiles of their spouse and two or more offspring. Category 4 scenarios involve pedigrees in which the genetic profile of a missing parent is compared with the reference genetic profiles of one or more offspring. In each category, the scenarios increase by a single child (sibling or offspring). Pedigrees were constructed for each scenario under all the possible combinations permitted by the type and quantity of individuals in a particular family. During the construction of each pedigree, STR profiles were imported onto the pedigree and a target individual was selected as the missing individual. Family pedigrees involving half siblings were constructed in a way in which the half sibling was considered to be a full sibling. Kinship analysis was performed to obtain a kinship index (KI) calculated by the LISA software for each population group (AFR-African American, CAU-Caucasian, and HIS-Hispanic) for every pedigree made and will be recorded in a Excel® spread sheet.

The dataset was checked twice to find any errors that can be corrected. A quick concordance study using eight random families was done using the validated spreadsheet DOJ VIEW 2.0. The study was conducted for the purpose of finding consistency between the KI values from pedigrees from LISA and DOJ VIEW 2.0 because both programs use Brenner's kinship algorithms [22, 31]. A quick reproducibility study was conducted to

see if LISA can reproduce the same KI values when kinship analysis is performed again on the same pedigrees from the eight families being studied. The results of the reproducibility study were compared to the KI values that were recorded at the beginning of the validation study.

The range and mean for the KI was determined for each scenario and population group using the statistical features of the Microsoft Excel® program. The KI dataset was converted into the natural logarithm (ln) of the KI to better observe the distribution of the data set graphically for each scenario. The number of observations was noted for each interval or range of the ln KI for each scenario and population group. To compare scenarios with other scenarios of the same category, the number of observations per ln KI range for each population group was combined. The validation study will not encompass kinship analysis of multigenerational families or families in which inbreeding is present. The study will only use STR data to calculate kinship indexes (KIs) even though LISA can calculate kinship indexes using Y-STR and mtDNA data. In all the validation study will focus on the efficacy of performing kinship analysis and pedigree searching using STR data.

As part of the validation study, any software problem encountered during the use of the LISA software and any recommendations were reported to FTI© to improve the overall performance of the program. An instructional protocol or manual was written up for the laboratory in how to effectively use the LISA software for kinship analysis which can be found in Appendix C on page 116.

CHAPTER V

RESULTS

Seven hundred and eighty nine STR profiles from 139 paternity cases were manually entered into LISA using the Profile Management. The Kinship Analysis module was used to build and analyzed an estimated that 15,800 pedigrees for 139 families assembled for this study based on the criteria for each of the four major categories of scenarios. A concordance check of KI values from pedigrees of eight random families interrogated with LISA and DOJ VIEW 2.0 found that half of the families produced consistent KI values between the two software packages. Many of the inconsistencies found were traced back to their investigation files in LISA and were attributed to software problems in LISA that corrupted the investigation files. The affected investigation files were remade and reanalyzed resulting in KI values consistent between the software packages. The reproducibility study found two families out of the eight produced KI values that were different to what was originally recorded on the Excel® spreadsheet. Again the inconsistency was traced back to the investigation files where the error was found. The affected investigation files were remade and reanalyzed resulting in KI values consistent with KI values that were originally recorded. The problems found by the concordance and reproducibility studies were reported to FTI© in order to be reviewed and corrected.

The mean and range of the KI for each scenario and population group are found in pages 51, 57, 61, and 66. The KI values were then converted into the natural logarithm (ln) KI in order to obtain the number of observations for each range of the ln KI for each scenario and population group. The results of this part of the study can be found in Appendix A on page 83. The number of observations for each population group for each scenario was combined and compared to other scenarios of the same category; the results for this part of the study are found in pages 52, 53, 58, 62, 63, 67, and 68.

Category 1

Category 1 Scenarios: Kinship Indexes (KI)						
Scenarios	AFR		CAU		HIS	
	Range	Mean	Range	Mean	Range	Mean
Scenario 1	1.20E+16 – 2.90E-25	2.05E+13	1.20E+17 – 2.80E-25	6.67E+13	3.30E+18 – 3.8E-25	1.38E+15
Scenario 2	2.30E+28 – 3.10E-30	8.83E+24	2.20E+30 – 7.00E-30	8.45E+26	2.20E+30 – 7.00E-30	8.45E+26
Scenario 3	2.60E+32 – 1.10E-24	1.78E+29	1.20E+32 – 5.60E-25	8.20E+28	1.60E+33 – 6.90E-25	1.09E+30
Scenario 4	1.20E+17 – 2.50E-23	3.81E+14	2.40E+16 – 4.70E-24	6.91E+13	2.30E+17 – 5.30E-24	5.12E+14
Scenario 5	7.40E+14 – 7.70E-23	9.66E+12	1.30E+14 – 1.10E-23	1.53E+12	4.60E+14 – 1.50E-23	4.91E+12
Scenario 6	2.50E+07 – 6.10E-14	2.84E+06	5.30E+06 – 5.20E-13	1.57E+06	6.80E+06 – 2.40E-13	1.24E+06

Table 3: The range and mean of the kinship index (KI) for each population group (AFR-African American, CAU-Caucasian, and HIS-Hispanic) and for every scenario from Category 1.

Category 1 scenarios simulate situations in which the DNA profile of a child is compared to the genetic data of a reference family pedigree with one parent and one or more siblings. Variation in KI values was observed for each pedigree analyzed for each

scenario and population group showing KI values as high 1.6E+32 and as low 7.0E-30. Pedigrees with KI values below 1 were attributed to pedigrees that consider a half sibling within a pedigree erroneously as a full sibling. The range and mean of KI values typically increase in subsequent scenarios but values begin to decrease after Scenario 3 because the number of large families that can produce reference pedigrees that contain one parent with four siblings are few and there are many more pedigrees that treat half siblings as full siblings. The trend observed with the range and mean of the KI was observed for all three population groups in Table 3.

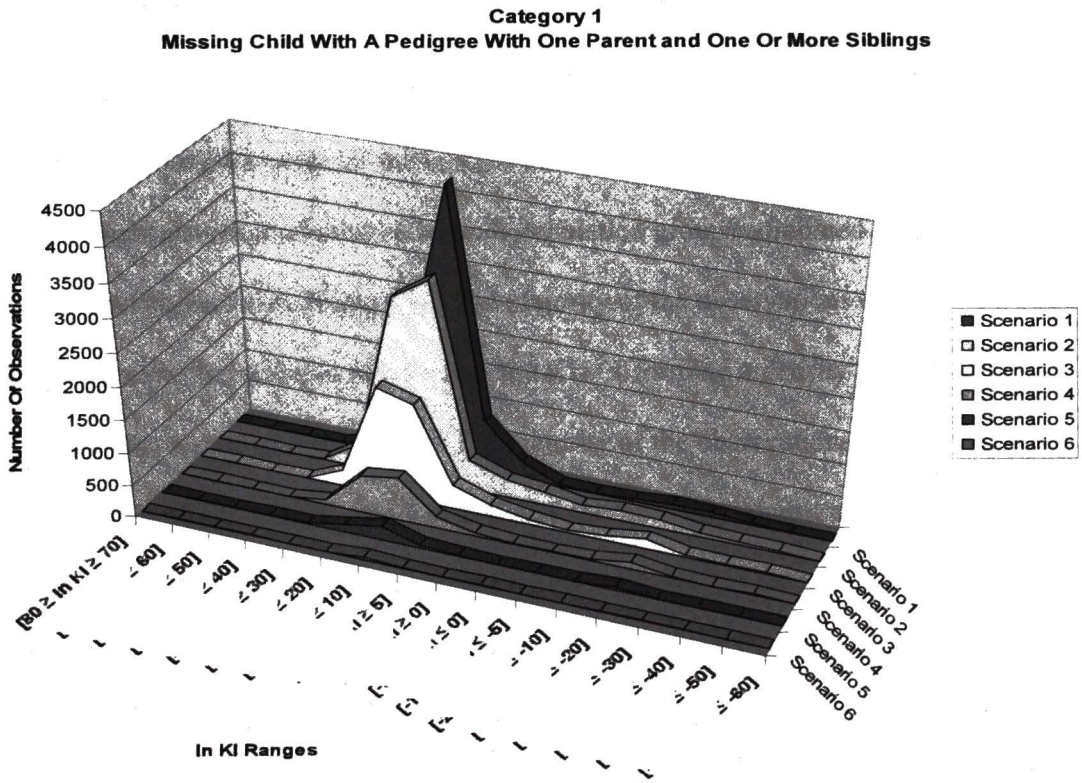


Figure 3: The distribution of the number of observations for all population groups combined versus the ranges of ln KI for each scenario in Category 1.

Category 1	Scenarios					
ln KI Ranges	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
[-70 ≤ ln KI ≤ -60]		4				
[-60 ≤ ln KI ≤ -50]	13	30	30	18	3	
[-50 ≤ ln KI ≤ -40]	31	52	20	0	0	
[-40 ≤ ln KI ≤ -30]	68	70	38	19	0	1
[-30 ≤ ln KI ≤ -20]	86	142	149	104	39	5
[-20 ≤ ln KI ≤ -10]	38	162	72	18	3	0
[-10 ≤ ln KI ≤ -5]	14	122	28	8	0	0
[-5 ≤ ln KI ≤ 0]	38	206	80	19	0	0
[5 ≥ ln KI ≥ 0]	274	277	189	50	4	0
[10 ≥ ln KI ≥ 5]	810	471	372	173	46	2
[20 ≥ ln KI ≥ 10]	4238	3112	1503	665	205	34
[30 ≥ ln KI ≥ 20]	1774	2757	1667	581	104	
[40 ≥ ln KI ≥ 30]	157	384	241	70	10	
[50 ≥ ln KI ≥ 40]	1	18	0	0		
[60 ≥ ln KI ≥ 50]		2	0			
[70 ≥ ln KI ≥ 60]		3	0			
[80 ≥ ln KI ≥ 70]			3			

Table 4: The number of observations for all population groups combined for each range of the ln KI and each scenario in Category 1.

The KI for each pedigree was converted into the ln KI, to obtain the number of observations for each range of the ln KI. A bimodal distribution of the number of observations per range of the ln KI for all population groups combined was observed in the log transformed data (Figure 3). The majority of the observations or pedigrees in all scenarios were found above ln KI = 0 or the KI = 1 forming a density curve that approximates a normal distribution. The majority of the observations were found in a range of 0 to 40 of the ln KI and 1 to 2.4E+17 for the KI for almost all of the scenarios in Category 1. The small distribution found below the ln KI = 0 for most scenarios in the range of 0 to -40 of the ln KI or 1 to 4.2E-18 for the KI. The small density curve consists of observations or pedigrees that consider a half sibling as a full sibling. The bimodal distribution trend found in Scenario 1 is repeated in subsequent scenarios when additional siblings are added to the reference pedigree.

Looking at the distribution of the ln KI for all possible pedigree combinations allowed in each scenario using two families from the dataset as an example (Families 00-0004 and 00-003) to see the trend observed in Category 1 scenarios. The pedigrees of Families 00-0004 and 00-0033 are two Hispanic families each with two parents and four offspring (Figure 4). Only Family 00-0033 contains a maternal half sibling (C1) amongst the four offspring.

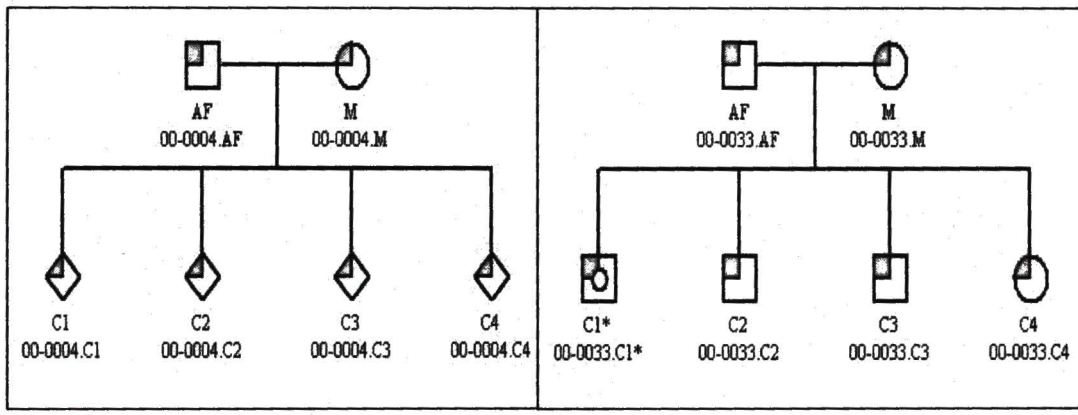


Figure 4: Family pedigree 00-0004 and 00-0033. C1* is the maternal half sibling.

Both families have pedigrees assignable to Scenarios 1, 2, and 3 for Category 1 because each family has four children and the maximum number of siblings in a reference pedigree is three which represents Scenario 3. A trend is observed in the distribution of the KI for all the possible pedigree combinations for each subsequent scenario (Figures 5 through 8).

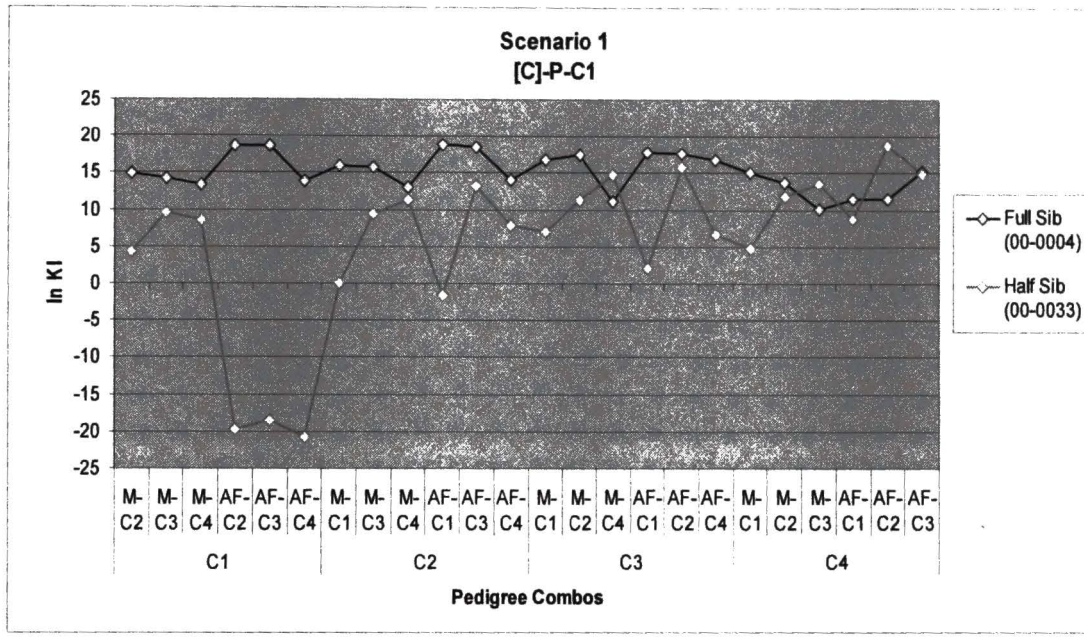


Figure 5: The ln KI distribution for pedigrees under Scenario 1 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child.

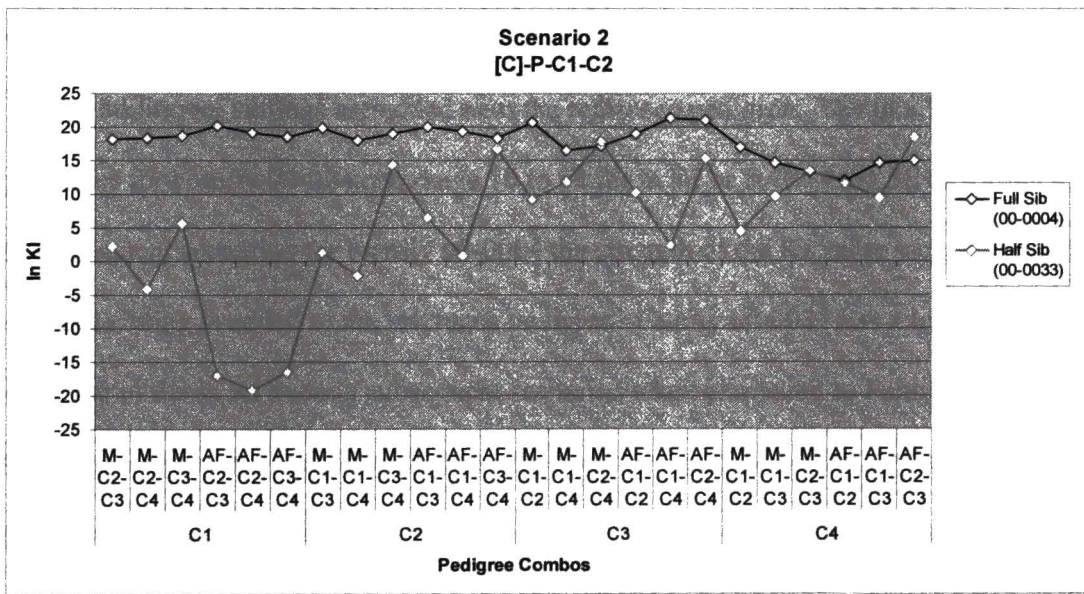


Figure 6: The ln KI distribution for pedigrees under Scenario 2 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child.

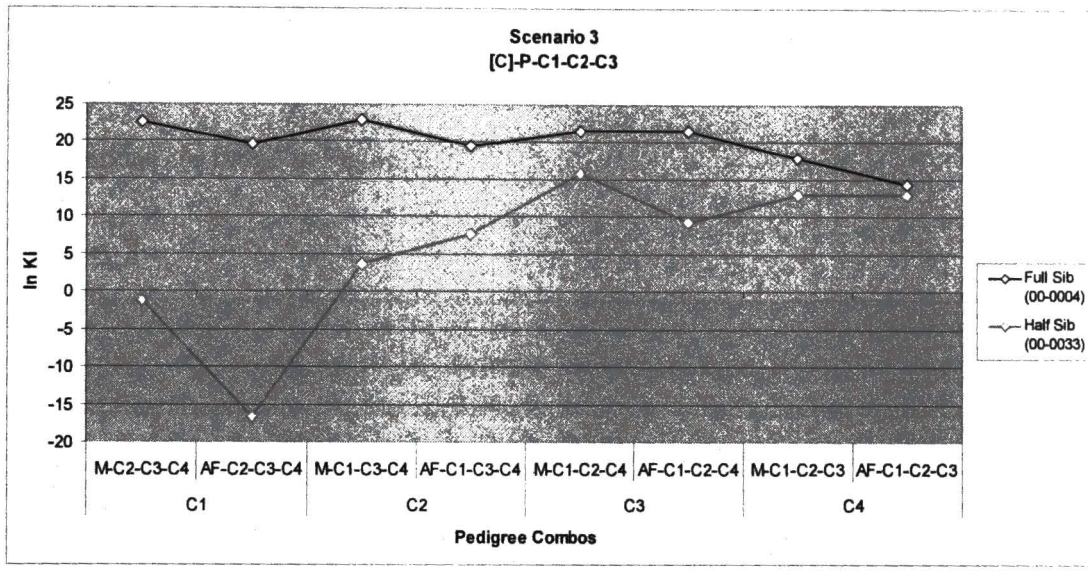


Figure 7: The ln KI distribution for pedigrees under Scenario 3 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child.

Variability in the KI was observed in each scenario for both families. The trend observed in Scenario 1 in Figure 5 is repeated again in subsequent scenarios of Category 1 when additional siblings are added onto the reference pedigree. The ln KI values for pedigrees for Family 00-004 can range 10 to 18 for ln KI or 2.4E+4 to1.5E+8 for the KI. In subsequent scenarios for Family 00-0004 the strength of the KI value increases when an extra sibling is added to the reference pedigree. Family 00-0033 has a lot more variation in the ln KI even in subsequent scenarios when compared to Family 00-0033. A trend can be seen in Family 00-0033 in which any pedigree involving the half sibling (C1) causes a drop in the KI value. The drop is even greater when the alleged father is part of the reference pedigree because the alleged father is not the biological father of the half sibling. If the correct pedigree was built in Family 00-0033 then the KI or ln KI

would have increased in value above KI = 1 or ln KI = 0. The ln KI values for the pedigrees in each scenario fall within the range ln KI where the large and small distributions are found in Figure 3.

Category 2

Category 2 Scenarios: Kinship Indexes (KI)						
Scenarios	AFR		CAU		HIS	
	Range	Mean	Range	Mean	Range	Mean
Scenario 7	1.50E+16 – 4.80-12	1.53E+13	1.10E+16 – 2.80E-13	1.20E+13	2.30E+16 – 4.10E-13	2.46E+13
Scenario 8	7.00E+15 – 3.60E-16	1.38E+13	6.40E+14 – 2.40E-16	2.89E+12	1.10E+15 – 6.90E-17	2.97E+12
Scenario 9	6.90E+34 – 1.20E-13	2.23E+32	8.80E+28 – 1.00E-14	2.84E+26	5.80E+26 – 8.30E-15	1.87E+24
Scenario 10	1.50E+14 – 5.40E-13	2.71E+12	8.20E+13 – 1.30E-13	1.19E+12	3.10E+14 – 3.40E-14	4.36E+12
Scenario 11	1.10E+08 – 5.10E-07	3.18E+07	1.90E+08 – 1.40E-06	6.64E+07	1.40E+08 – 2.80E-07	2.88E+07

Table 5: The range and mean of the kinship index (KI) for each population group (AFR-African American, CAU-Caucasian, and HIS-Hispanic) and for every scenario in Category 2.

Category 2 scenarios simulate situations in which the DNA profile of a missing child is compared to the genetic data of reference family pedigree that consists of two or more siblings. The mean and range of the KI increases but decreases after Scenario 9 just like Category 1 because the number of families that can have references pedigrees with five or more siblings are few and most of these families have erroneous pedigrees with half siblings treated as full siblings. The values for the KI range and the mean are much lower compared to Category 1 because the parents are not part of the reference pedigree. Variation in the KI was observed for many pedigrees ranging from 6.90E+34 to 6.90E-17.

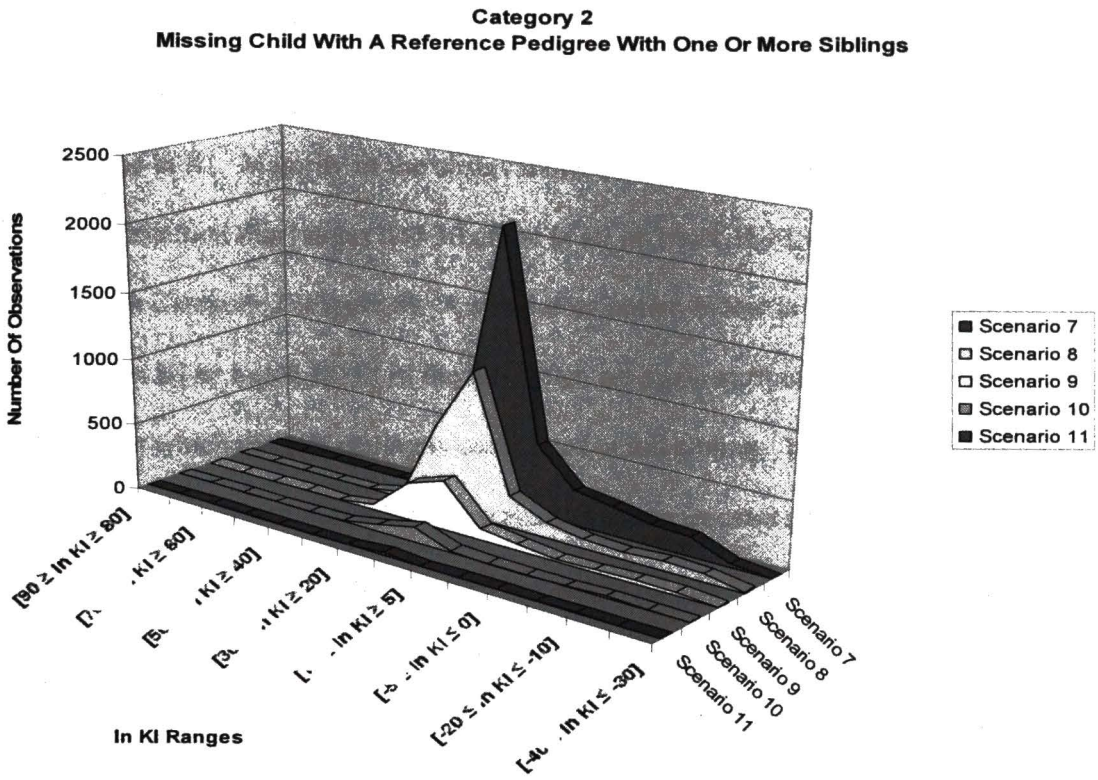


Figure 8: The distribution of the number of observations for all population groups combined versus the ranges of ln KI for each scenario in Category 2.

Category 2	Scenarios				
ln KI Ranges	Scenario 7	Scenario 8	Scenario 9	Scenario 10	Scenario 11
[-40 ≤ ln KI ≤ -30]		7	6	1	
[-30 ≤ ln KI ≤ -20]	33	59	27	5	
[-20 ≤ ln KI ≤ -10]	150	68	30	15	3
[-10 ≤ ln KI ≤ -5]	160	52	25	3	
[-5 ≤ ln KI ≤ 0]	218	65	19		
[5 ≥ ln KI ≥ 0]	264	112	51	3	
[10 ≥ ln KI ≥ 5]	538	239	88	16	
[20 ≥ ln KI ≥ 10]	2097	1102	369	114	18
[30 ≥ ln KI ≥ 20]	700	635	277	55	
[40 ≥ ln KI ≥ 30]	49	49	35	4	
[50 ≥ ln KI ≥ 40]					
[60 ≥ ln KI ≥ 50]					
[70 ≥ ln KI ≥ 60]			2		
[80 ≥ ln KI ≥ 70]					
[90 ≥ ln KI ≥ 80]			1		

Table 6: The number of observations for all population groups combined for each range of the ln KI and each scenario in Category 2.

A bimodal distribution trend similar to Category 1 distribution was observed when the KI is converted to the ln KI and the number of observations is counted for each range of the ln KI. Majority of the observations are found above ln KI = 1 or KI = 0 with a large distribution that ranges 0 to 30 of the ln KI or 1 to 1.1E+13 for the KI in most scenarios in Category 2 (Figure 8). A small distribution below ln KI = 0 ranged from 0 to -30 of the ln KI or 1 to 9.4E-14 of the KI which are again attributed to erroneous pedigrees in which the half sibling is considered a full sibling. The trend observed in Scenario 7 can be found in subsequent scenarios where an extra sibling is added to the reference pedigree. The trends observed in Category 2 can be better understood by looking again at Families 00-004 and 00-0033.

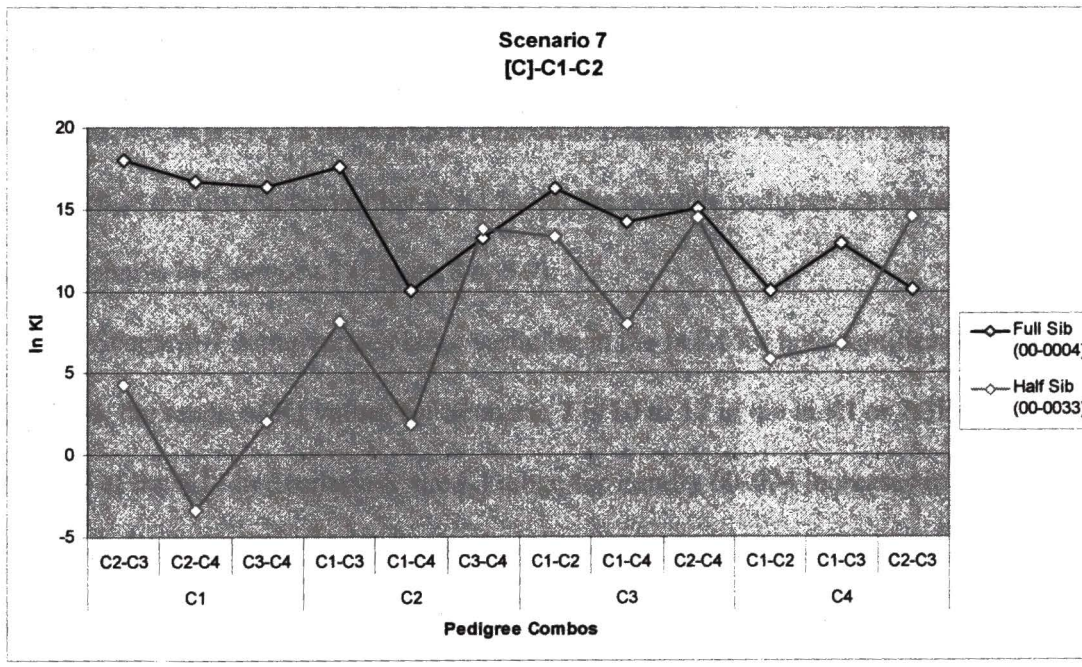


Figure 9: The ln KI distribution for pedigrees under Scenario 7 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child.

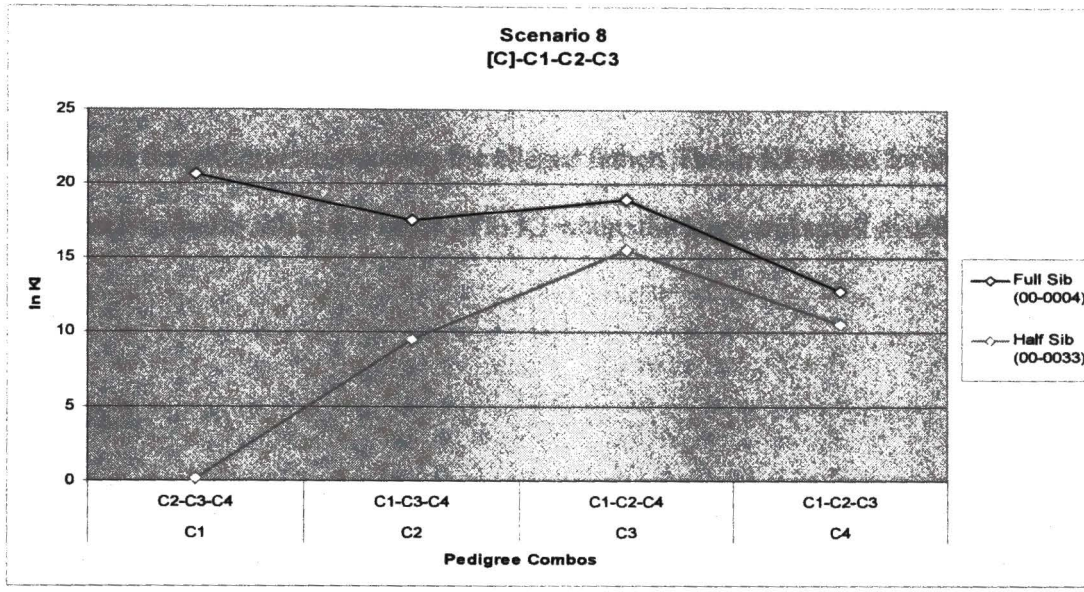


Figure 10: The ln KI distribution for pedigrees under Scenario 8 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child.

Both families have pedigrees in Scenario 7 and 8 because each family has four children meaning that the maximum number of siblings that can be added to a reference pedigree in Category 2 is three which represents Scenario 8. A common trend is observed in the distribution of the ln KI for all the possible pedigree combinations for each scenario in Category 2 (Figure 8 and 9).

Scenario 7 demonstrates large variation in the ln KI for both families. In Family 00-0004, the range of KI values for Scenario 7 is 10 to 17 in the ln KI or 2.3E+4 to 6.5E+7 of the KI. In Scenario 8, the KI range for Family 00-004 increases in value 2 to 20 of the ln KI or 3.5E+5 to 9.40E+8 of the KI. Family 00-0033 shows a lot more variability in the KI for both scenarios compared to Family 00-0004. The strength of the ln KI values for Family 00-0033 actually increases in value when an additional sibling is added to the reference pedigree in the subsequent scenario. Again like in Category 1

scenarios, pedigrees containing the half sibling (C1) have ln KI values below 0 and it is even worst for pedigrees containing the alleged father. The ln KI values for the pedigrees under each scenario fall in the range of ln KI where the large and small distributions are found.

Category 3

Category 3 Scenarios: Kinship Indexes (KI)						
Scenarios	AFR		CAU		HIS	
	Range	Mean	Range	Mean	Range	Mean
Scenario 12	1.9E+12 – 2.60E-33	9.87E+09	6.60E+14 – 4.00E-33	8.89E+11	4.10E+15 – 2.70E-33	4.77E+12
Scenario 13	9.60E+20 – 1.50E-33	1.02E+18	1.70E+23 – 2.50E-33	1.46E+20	4.80E+23 – 1.60E-33	4.11E+20
Scenario 14	2.50E+21 – 1.50E-20	1.07E+19	1.80E+22 – 2.30E-20	2.85E+19	3.20E+22 – 5.90E-21	4.68E+19
Scenario 15	7.10E+21 – 7.20E-16	1.18E+20	1.60E+24 – 1.10E-15	5.17E+21	4.20E+24 – 2.80E-16	1.36E+22
Scenario 16	1.10E+22 – 3.60E-13	2.15E+20	6.60E+20 – 5.40E-13	1.23E+19	4.10E+21 – 1.40E-13	6.61E+19
Scenario 17	7.60E+18 -1.80E-10	3.71E+17	4.70E+17 – 2.70E-10	2.94E+16	6.30E+16 – 7.00E-11	4.41E+15
Scenario 18	3.90E+00 – 3.00E-02	1.97E+00	2.50E+00 – 6.60E-02	1.28E+00	5.20E+00 – 4.60E-02	2.62E+00

Table 7: The table shows the range and mean of the kinship index (KI) for each population group (AFR-African American, CAU-Caucasian, and HIS-Hispanic) and for every scenario in Category 3.

Category 3 scenarios simulate situations in which the DNA profile of a missing parent is compared to the genetic data of a reference family pedigree that consists of the spouse and one or more offspring. Variation in the KI was observed for each pedigree analyzed for each scenario and race with KI values as high 4.20E+24 and as low as 2.70E-33. Like in previous scenarios from Categories 1 and 2, an increase in the mean

and range of the KI is observed in subsequent scenarios until after Scenario 16 has been reached where families that have reference pedigrees with five or more offspring are few and contain a half sibling.

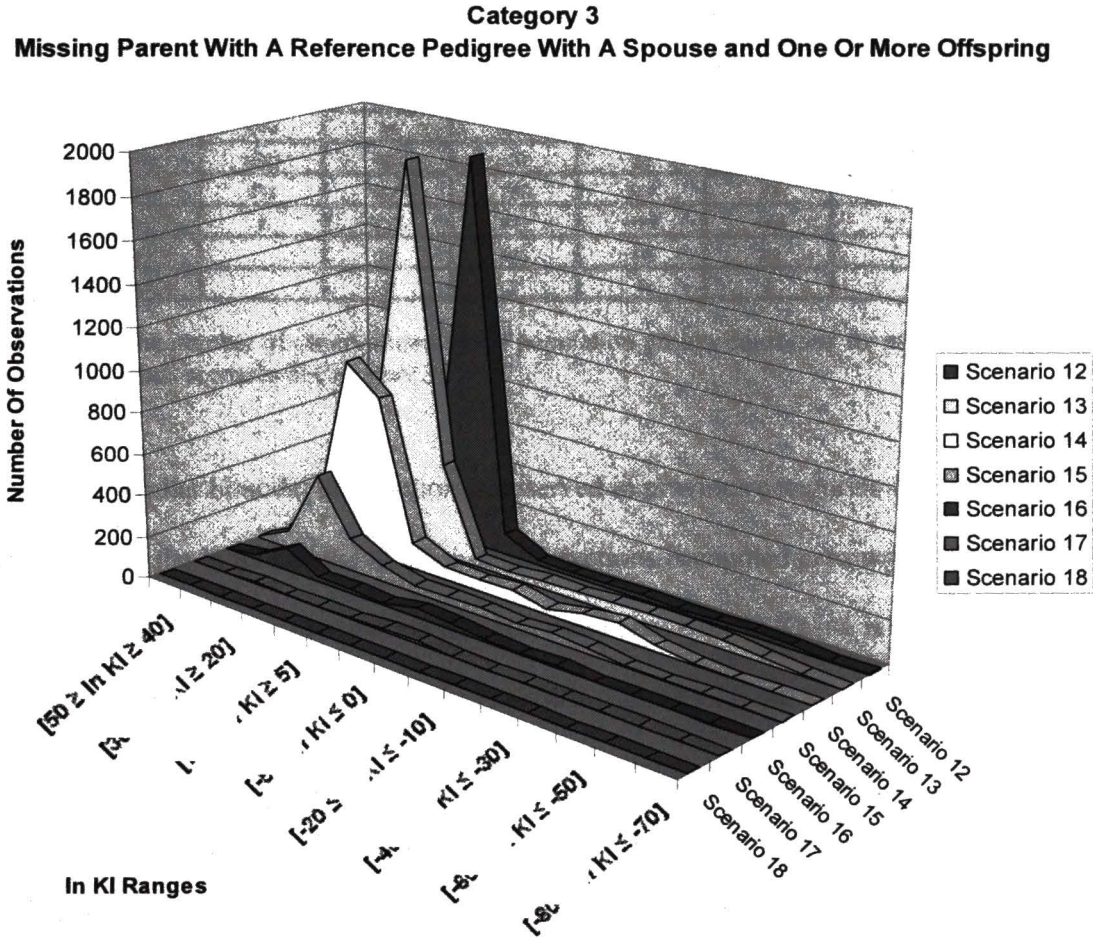


Figure 11: The distribution of the number of observations for all population groups combined versus the ranges of ln KI for each scenario in Category 3.

Category 3	Scenarios						
ln KI Ranges	Scenario 12	Scenario 13	Scenario 14	Scenario 15	Scenario 16	Scenario 17	Scenario 18
[-80 ≤ ln KI ≤ -70]	3	3					
[-70 ≤ ln KI ≤ -60]	3	9					
[-60 ≤ ln KI ≤ -50]	17	30					
[-50 ≤ ln KI ≤ -40]	25	51	6				
[-40 ≤ ln KI ≤ -30]	18	57	31	4			
[-30 ≤ ln KI ≤ -20]	15	58	78	25	15	3	
[-20 ≤ ln KI ≤ -10]	23	53	60	38	7		
[-10 ≤ ln KI ≤ -5]	21	42	27	23	15	3	
[-5 ≤ ln KI ≤ 0]	15	45	63	39	31	15	3
[0 ≥ ln KI ≥ 0]	28	32	53	39	45	18	3
[10 ≥ ln KI ≥ 5]	106	36	59	34	14	4	
[20 ≥ ln KI ≥ 10]	1883	455	127	74	25	4	
[30 ≥ ln KI ≥ 20]	433	1875	786	150	26	4	
[40 ≥ ln KI ≥ 30]	2	710	921	413	114	12	
[50 ≥ ln KI ≥ 40]		52	181	86	24	3	
[60 ≥ ln KI ≥ 50]		2	2	5	2		

Table 8: The number of observations for all population groups combined for each range of the ln KI and each scenario in Category 3.

A similar bimodal distribution is observed in previous categories when the KI is converted to the ln KI and the number of observations is counted per range of the ln KI (Figure 11). The major distribution that is formed above the ln KI = 1 is located at a range between 10 to 40 of the ln KI or 2.2E+4 to 2.3E+17 for the KI for most scenarios in Category 3. As for the small distribution observed below ln KI = 0 ranges between 0 to -30 of the ln KI which is 1 to 9.4E-14 in KI. Again the trend repeats itself when an extra offspring is added to the reference pedigree in subsequent scenarios. To understand better the observed trend is by looking at the distribution of the ln KI for each pedigree combination for each scenario of Category 3 for Families 00-0004 and 00-0033 (Figures 12 through 15).

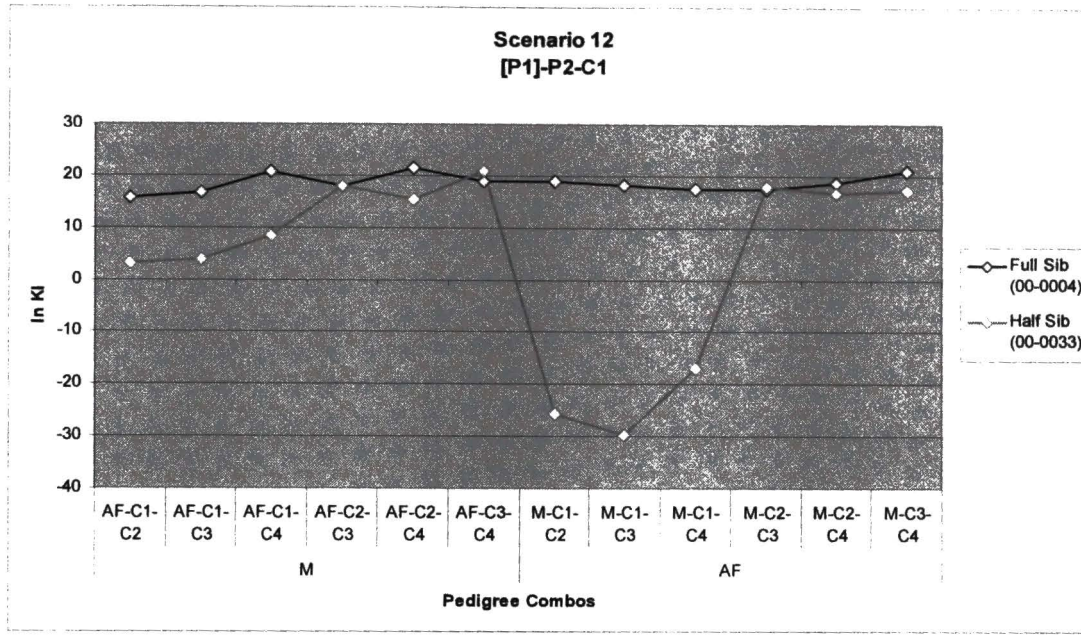


Figure 12: The ln KI distribution for pedigrees under Scenario 12 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

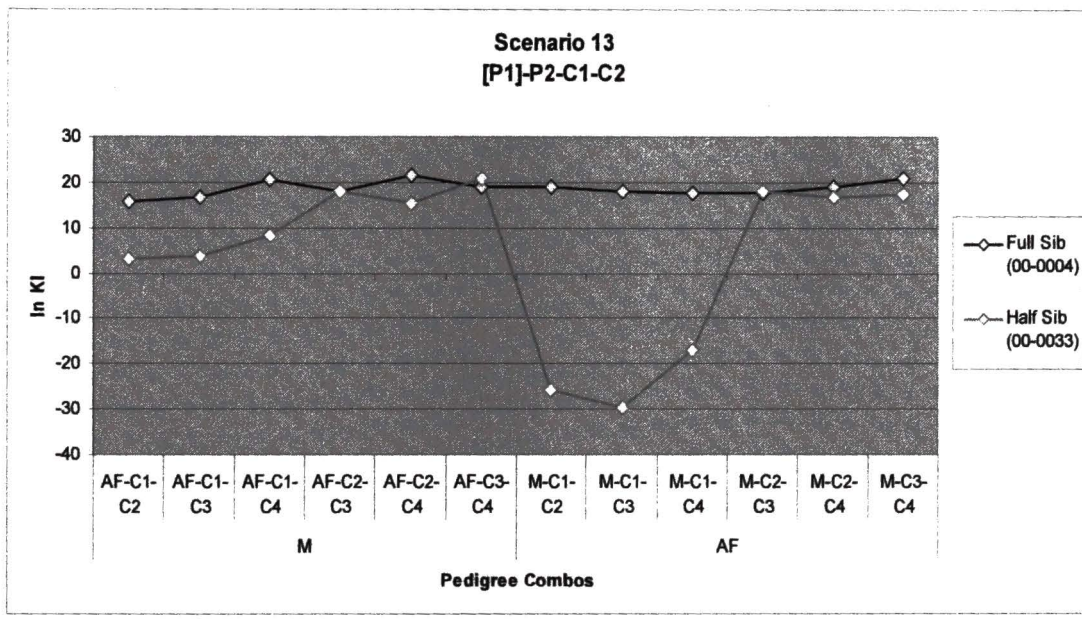


Figure 13: The ln KI distribution for pedigrees under Scenario 13 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

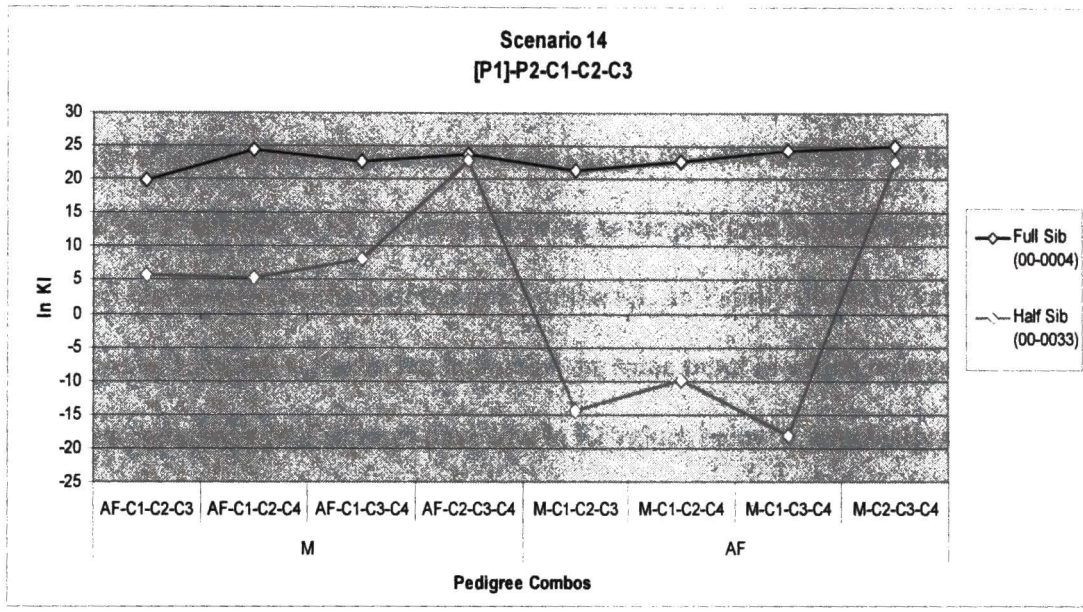


Figure 14: The ln KI distribution for pedigrees under Scenario 14 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

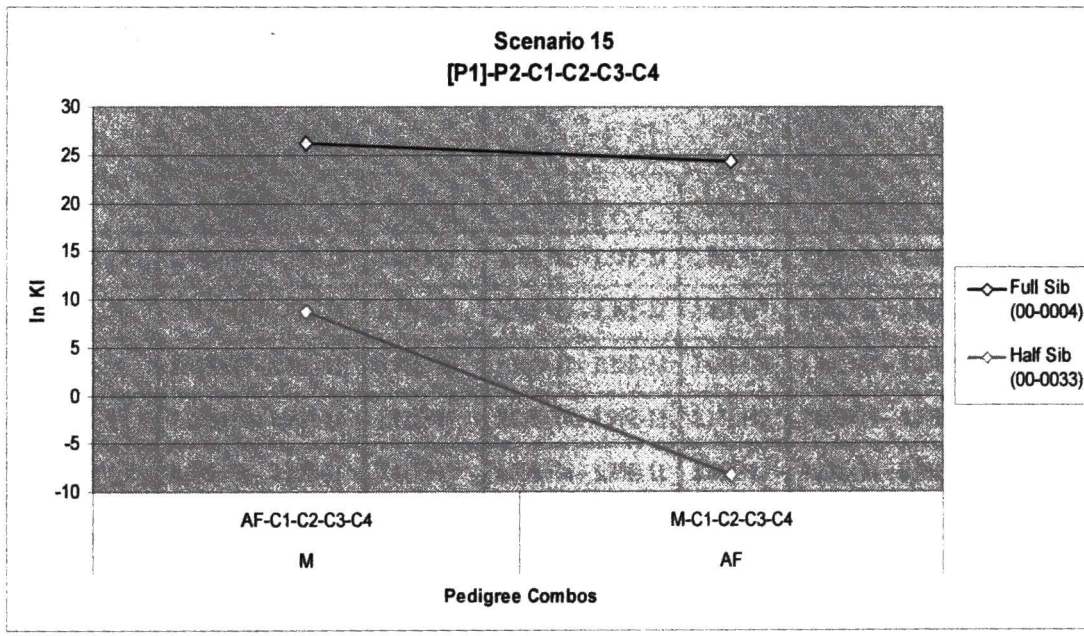


Figure 15: The ln KI distribution for pedigrees under Scenario 15 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

Both families have pedigrees between Scenarios 12 to 15 because the maximum number offspring that can be added to a reference pedigree is four which is Scenario 15. The range of ln KI values for Scenario 12 for Family 00-0004 is 8 to13 for the ln KI or 3.50E+05 to 9.40E+08 in KI. Adding offspring to the pedigree in subsequent scenarios causes an increase in the value of the ln KI or the KI. In Family 00-0033, variability ln KI is observed in each scenario but maintains the same trend in which reference pedigrees containing half siblings have low ln KI values below 0, especially when the alleged father is the missing person. The ln KI values for pedigrees in Family 00-033 fall within the ranges where the large and small distributions are found.

Category 4

Category 4 Scenarios: Kinship Indexes (KI)						
Scenarios	AFR		CAU		HIS	
	Range	Mean	Range	Mean	Range	Mean
Scenario 19	1.40E+17 – 1.30E-23	1.47E+14	3.50E+19 – 1.50E-23	2.80E+16	1.10E+20 – 8.50E-24	8.88E+16
Scenario 20	1.20E+18 – 4.40E-22	2.79E+15	3.50E+20 – 3.50E-22	6.45E+17	1.00E+21 – 3.90E-22	2.74E+18
Scenario 21	8.70E+37 – 8.30E-17	3.01E+35	1.30E+32 – 1.20E-16	7.68E+29	1.40E+29 – 3.20E-17	7.22E+26
Scenario 22	1.20E+19 – 1.30E-14	1.12E+17	3.90E+16 – 1.90E-13	1.11E+15	1.50E+17 – 4.90E-14	2.70E+15
Scenario 23	6.00E+18 – 2.20E-11	2.61E+17	6.80E+14 – 3.20E-11	3.68E+13	9.20E+14 – 8.50E-12	4.12E+13
Scenario 24	6.10E+00 – 3.80E-03	3.05E+00	4.70E+00 – 5.60E-02	2.38E+00	7.00E+00 – 2.00E-02	3.51E+00

Table 9: The range and mean of the kinship index (KI) for each population group (AFR-African American, CAU-Caucasian, and HIS-Hispanic) and for every scenario in Category 4.

Category 4 simulates situations in which the DNA profile of the missing parent is compared to the genetic data from a reference family pedigree with two or more offspring. Variability of the KI values was observed for each pedigrees analyzed for each scenario ranging from 8.70E+37 to 8.50E-24. The range and mean of the KI increases when extra offspring is added to the reference family pedigree until Scenario 22 is reached. In Scenario 22, the amount of families that have references pedigrees that contain five or more offspring are few and the number erroneous pedigrees with a half sibling increases.

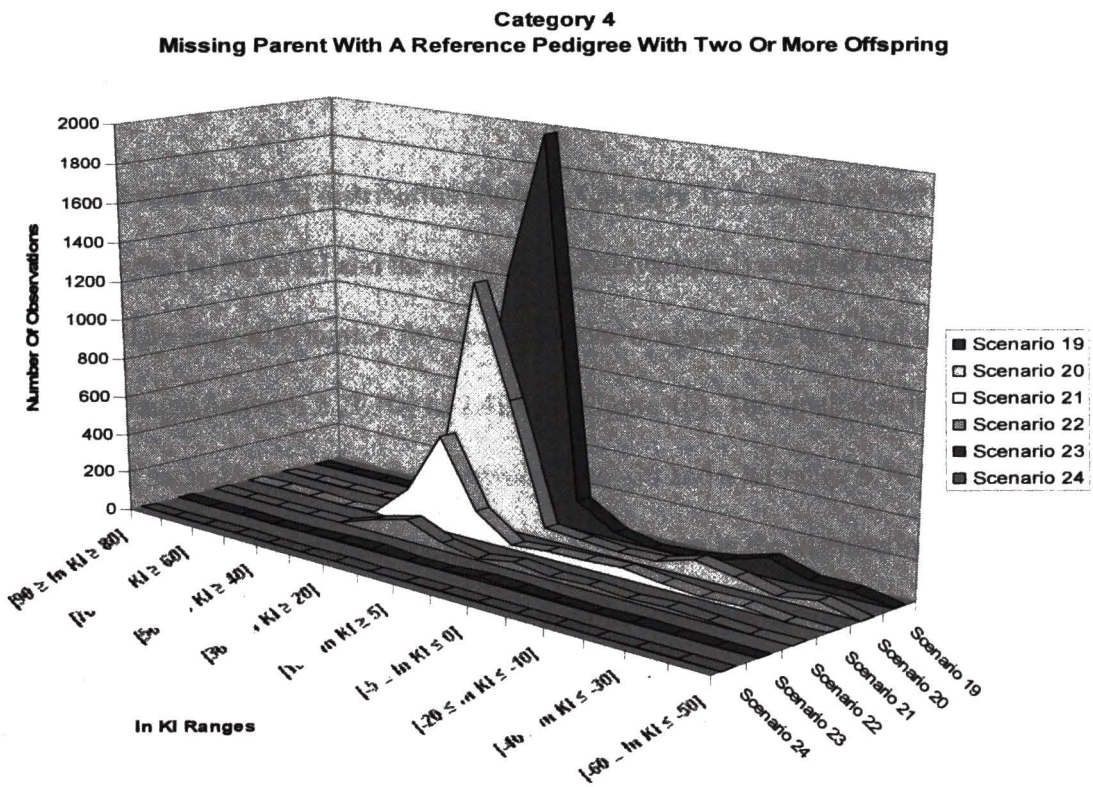


Figure 16: The distribution of the number of observations for all population groups combined versus the ranges of ln KI for each scenario in Category 4.

Category 4	Scenarios					
ln KI Ranges	Scenario 19	Scenario 20	Scenario 21	Scenario 22	Scenario 23	Scenario 24
[-60 ≤ ln KI ≤ -50]	4					
[-50 ≤ ln KI ≤ -40]	37	47				
[-40 ≤ ln KI ≤ -30]	39	16	24	8		
[-30 ≤ ln KI ≤ -20]	103	59	19	21	3	
[-20 ≤ ln KI ≤ -10]	42	95	32	10	1	
[-10 ≤ ln KI ≤ -5]	10	37	58	25	7	1
[-5 ≤ ln KI ≤ 0]	3	34	45	18	11	2
[5 ≥ ln KI ≥ 0]	34	46	43	37	18	3
[10 ≥ ln KI ≥ 5]	152	63	26	9	4	
[20 ≥ ln KI ≥ 10]	1985	678	155	40	3	
[30 ≥ ln KI ≥ 20]	1249	1243	497	110	11	
[40 ≥ ln KI ≥ 30]	109	270	174	65	10	
[50 ≥ ln KI ≥ 40]	4	13	10	2	1	
[60 ≥ ln KI ≥ 50]			1			
[70 ≥ ln KI ≥ 60]			4			
[80 ≥ ln KI ≥ 70]			5			
[90 ≥ ln KI ≥ 80]			2			

Table 10: The number of observations for all population groups combined for each range of the ln KI and each scenario in Category 4.

The same bimodal distribution found in Category 1, 2, and 3 is observed when the KI is converted to the ln KI and the number of observations is counted for each range of the ln KI. The large distribution for most scenarios occupies a range of 10 to 40 of the ln KI or a KI range between 2.2E+04 to 2.4E+17 (Figure 16). The small distribution for all scenarios, it occupies an ln KI range between -5 to -30 and a KI range 6.7E-3 to 9.3E-14 (Figure 16). The trend observed in Category 4 can be seen in the distribution ln KI for all pedigree combinations for each scenario for Families 00-0004 and 00-0033 (Figures 17 through 19).

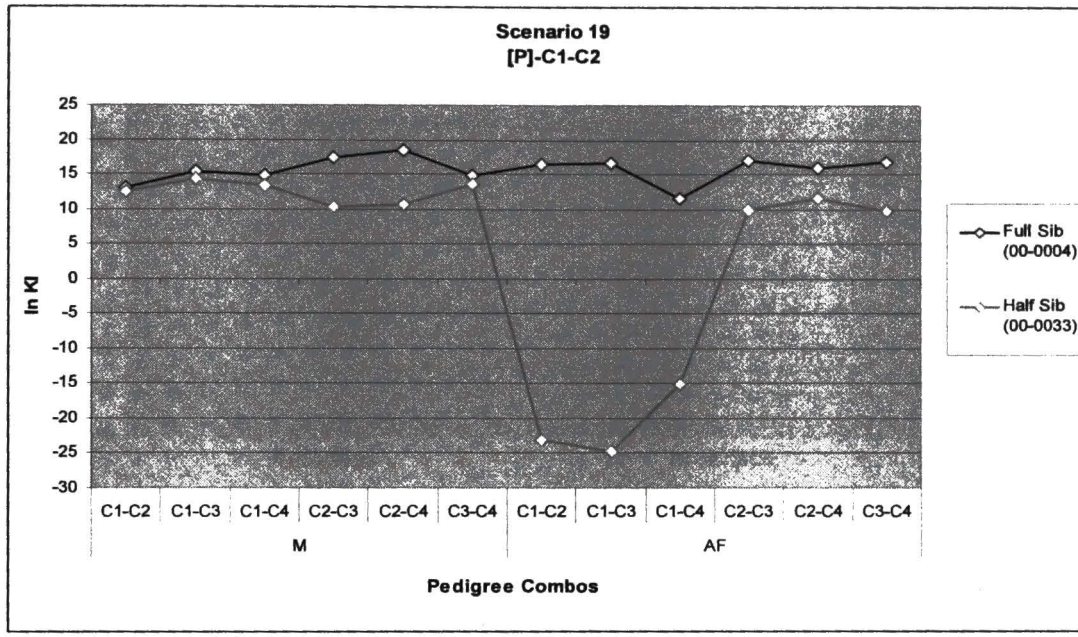


Figure 17: The ln KI distribution for pedigrees under Scenario 19 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

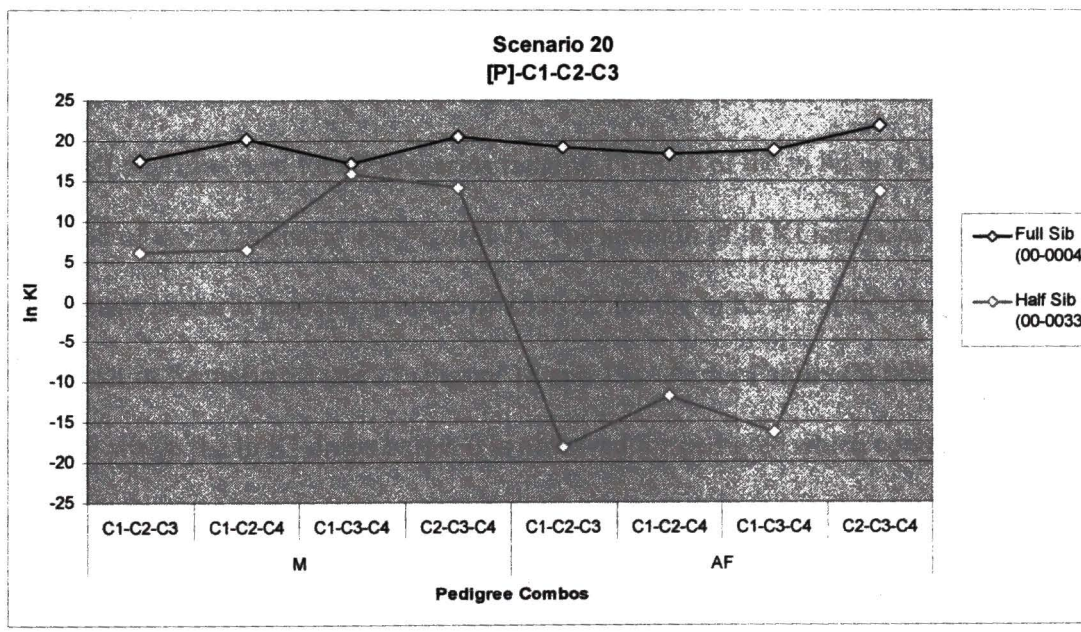


Figure 18: The ln KI distribution for pedigrees under Scenario 20 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

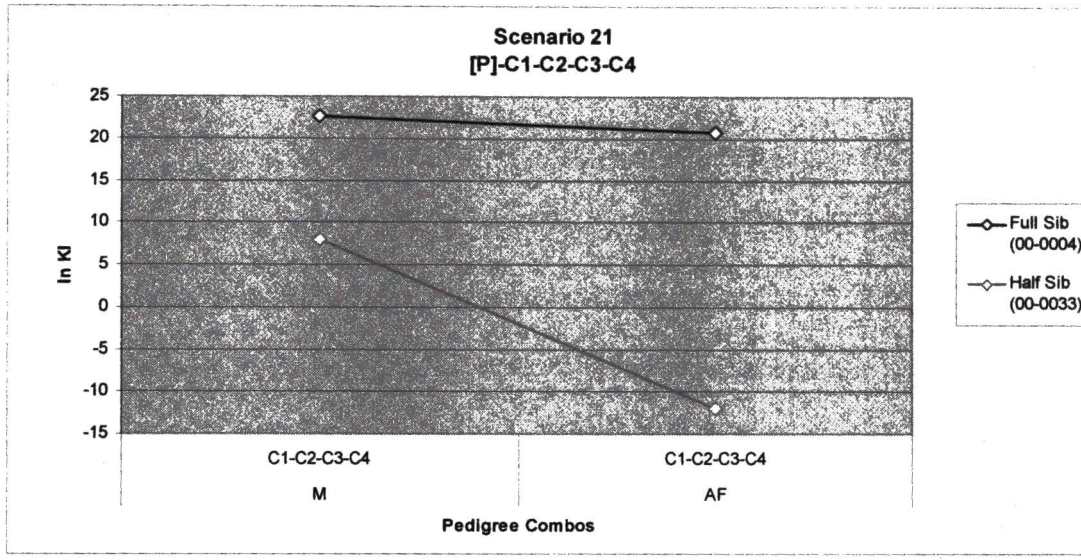


Figure 19: The ln KI distribution for pedigrees under Scenario 21 for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized M or AF is the missing or test parent.

Both families have pedigrees under Scenarios 19 through 21 for Category 4 because the family has four children meaning the maximum number of offspring in a reference pedigree is four which represents Scenario 19. In Family 00-0004, variation of the ln KI was observed for all scenarios ranging 11 to 18 of the ln KI or 1.10E+15 to1E+08 of the KI Scenario 19 (Figure17). The strength of ln KI increases for every subsequent scenario reaching a range of 20 to 22 for the ln KI or 1.10E+19 to 6.30E+9 for the KI of Scenario 20 and 21 (Figure 18 and 19). As for Family 00-0033 the same trend in which the ln KI drops below 0 is observed for pedigree which contain the half sibling (C1) and which the alleged father is the missing individual but ln KI values tend decrease or fluctuate in each successive scenario. For example, in Scenario 19 the range of ln KI values is -24 to 14 or a KI range 1.90E-11 to 1.70E+06 while Scenario 21 the range of the ln KI is -11 to 8 or 6.20E-6 to 3E+3 in Scenario 21. To view more KI and ln

KI values for all scenarios for Families 00-0004 and 00-0033 refer to Appendix B on page 108.

CHAPTER VI

CONCLUSION AND DISCUSSION

All four categories of scenarios demonstrated a general trend of a bimodal distribution in which the majority of the observations or pedigrees have KIs greater than one while other observations or pedigrees in which the half sibling is treated as a full sibling have KIs below 1. The major and minor density curves approximate a normal curve and the bimodal distribution is repeated in subsequent scenarios where children (siblings or offspring) were added successively to the pedigree. A similar bimodal distribution is observed when the number observations per range of the ln KI for all population groups and scenarios were combined for each major category of scenarios (Figure 20).

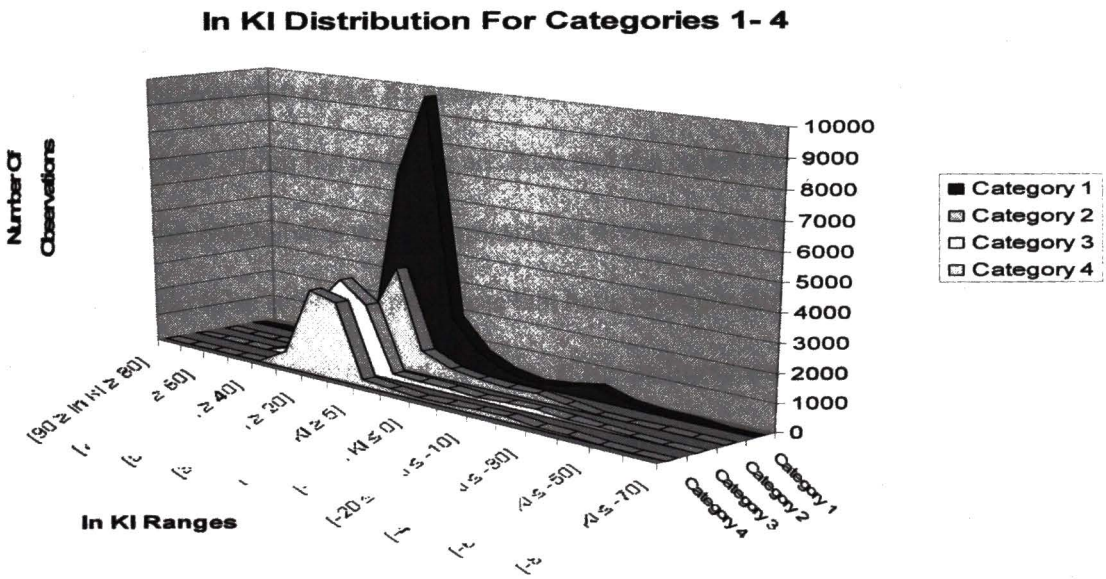


Figure 20: The distribution of the number of observations for all population groups and scenarios combined versus the ranges of ln KI for each category of scenarios.

Looking at the range and mean of the KI for each scenario, it was clear that the KI values increased and then decreased in value when subsequent children were added to the pedigree. Each category of the KI range and mean decreased at a particular scenario which was attributed to the small amount of large families that have five or more children and pedigrees with a single half sibling. Usually by adding another individual to a reference pedigree, the KI for the association usually increases in value but the type of relationship the individual has to the reference pedigree influences the out come of the KI [8]. For example, KI values were higher in Category 1 and 3 compared to Category 2 because a parent or spouse was part of the reference pedigree. However, by looking at the distribution of ln KI for all possible pedigrees for each scenario for one Family (00-0004), it was apparent that the ln KI values were increasing in value when additional siblings or offspring were added to the reference pedigree in subsequent scenarios.

The decrease in the KI range and mean in later scenarios of each category can also be attributed to families with a single half sibling that was treated as full sibling that caused the depreciation of the KI value. Any KI value below 1 is a good indicator that the relationships within the pedigree are false and the culprit could be a single individual that has a false relationship with the missing or deceased individual. The KI depreciated even in pseudo false reference pedigrees that have multiple siblings or offspring that have true relationships with parents and siblings. If the pedigree was built correctly in LISA by making the false full sibling as an actual half sibling the KI increases in value above 1 falling in the range of the large curve of the bimodal distribution. In Figure 20, the ln KI distribution of Scenario 1 for all possible pedigree combinations demonstrates that when

Family 00-0033 pedigrees are built correctly to account for the half sibling, the depreciated ln KI values rise in value above the ln KI = 0.

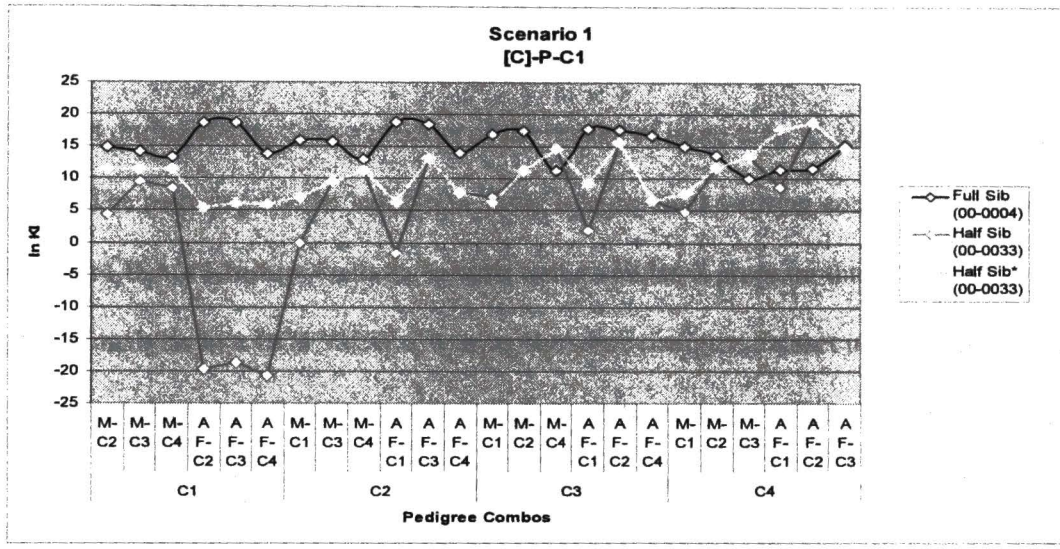


Figure 21: The ln KI distribution for pedigrees under Scenario1for families 00-0004 and 00-0033. M-mother, AF –alleged father, and C- child. Capitalized C is the missing or test child. The true family pedigrees for 00-0033 is the aqua line and the false the red line.

Another observation is the amount of variation seen in each family as well as the entire data set for each scenario. One factor for this observation is due to the biological relationships within the family that influences the calculations for the KI. Such factors can consist of mutations, missing loci, and biological relationships with in the family. Another factor is how the calculation for the KI is made, is the software using a particular method to calculate the KI (general or Brenner's method), rounding, minimum allele frequency settings, or the type of allele frequency tables used. Variability was observed when different allele frequency tables from any of the three population groups used in the study [3, 8]. A third major factor that must be accounted for the variability of the KI

within each family and the entire data set is error. The concordance and reproducibility study that was made showed that 4 out 8 families had inconsistent KI values compared to DOJ VIEW 2.0 results. Many of these inconsistencies with the KI results were attributed to problems in the software. Another type of error that cannot be discounted is errors that occurred during the recording process of the results of more than 15,800 pedigrees.

The next step for the validation study is to conduct a strong concordance study between LISA's Kinship Analysis module and a validated program that can perform kinship analysis like DOJ VIEW 2.0 using the same pedigree dataset used in this study. Looking at only eight families is not enough to make sure that results obtained by LISA are error free. It is also the only way to identify any errors in the dataset that can be attributed to problems within the software or errors caused by the recording process. A limitation arises when using DOJ VIEW 2.0 for a concordance study because the pedigrees in which the half sibling is treated as a full sibling cannot be analyzed causing the KI for such a pedigree in DOJ VIEW 2.0 always resulting in zero unlike LISA [22, 31]. Using another comparable kinship analysis program that has been validated by the lab can help avoid this limitation.

Next a validation study of the Searching module is needed because it's a critical component for a LIMS software that will be used for large scale investigations. The Searching module in LISA allows the analyst or investigator to make a pedigree search using the DNA profile of an unknown deceased individual in order to find an association or hit to particular reference family. The search generates a list of hits or associations that were made with calculated KIs for each population group [23, 31].

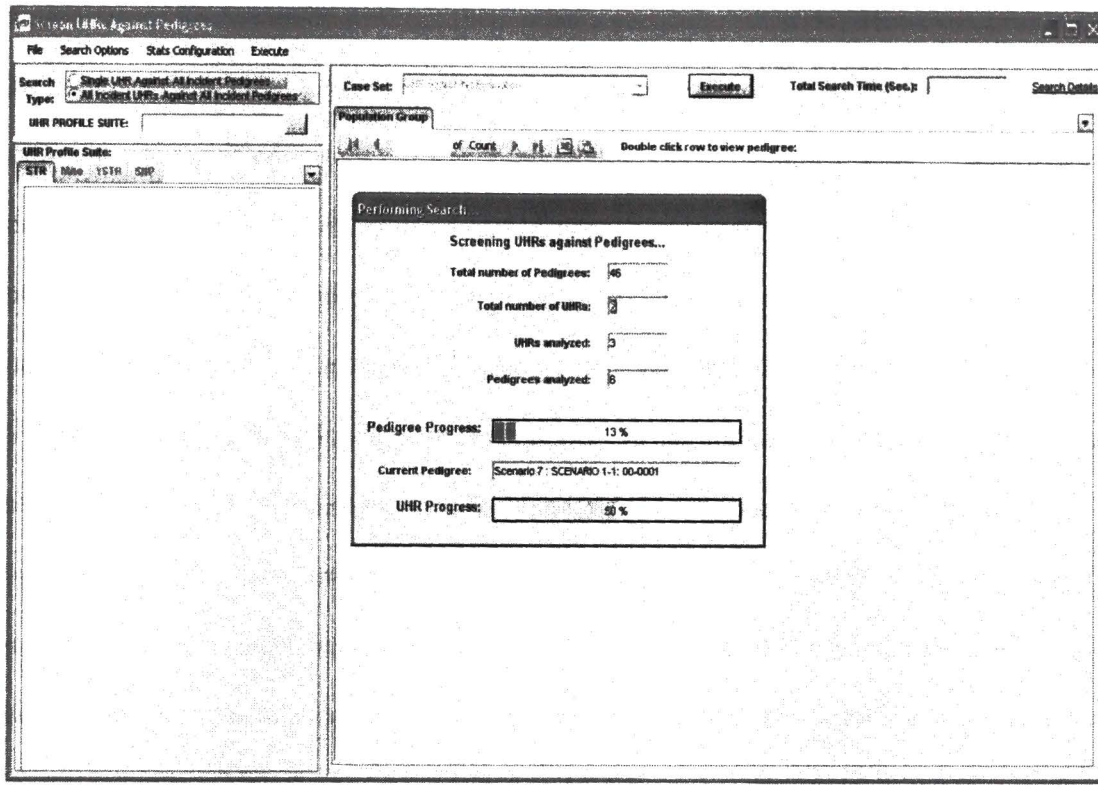


Figure 22: Screenshot of the Searching module.

In any search there will be a number of hits or associations that are false which also known as fortuitous kinship associations (FKAs). To measure the amount of FKAs for any of the four categories mentioned in this study is by using the DNA profiles from the paternity cases, duplicate the profiles using the cloning feature of the Profile Management module, and classify them as unidentified human remains (UHRs). A pedigree can be made using Search UHRs Against Pedigree feature of the searching module. Record the number of false hits the UHR profile makes with the incorrect pedigree that was made under a particular scenario. Convert the KI for the false hits into the ln KI and compare them to the ln KI ranges made in this study to find an overlapping

region where FKAs can be distinguished under a particular scenario. However, problems have arisen using LISA's Searching module making the validation study for that module difficult. One problem is that the Searching module cannot handle a pedigree search in the thousands because of the DNA VIEW™ portion of LISA. Every time a search is made, a hit is made between a UHR profile and a reference pedigree causing LISA to upload DNA VIEW™ to make the KI calculation. The problem occurs when DNA VIEW™ is constantly uploading for several hits causing the software to crash yielding no search results. FTI© has installed mtDNA and Y-STR filters to screen pedigrees with or without data from these markers and recommended a limit on the number of pedigrees made per case set to get around the problem. The problem with the Searching module is detrimental because there will be situations in which a lab will encounter a large scale investigation in which the number of dead and family references range in the thousands or in the hundred thousands like in the East Asian Tsunami of 2004 [4, 19, 31].

Validating Kinship Analysis and Searching modules of LISA is only the beginning of the validation of this LIMS software. Other features in the Kinship Analysis modules must be validated like Two Hypothesis feature which is used to calculate the likelihood ratio when two pedigrees are compared. Both the Kinship Analysis and Searching modules must be validated for mtDNA, Y-STR, and SNP data if possible. Other modules in the Statistical Application of LISA that needs to be validated are the Statistical Analysis and the Mixture Analysis modules. Any validation involving Statistical Analysis and the Mixture Analysis modules must include a strong concordance study using a validated analytical program. The concordance study is necessary in the

validation study of any module in LISA because it helps identify problems related to the software or recording errors [31].

LISA fulfills many of the recommendations for a LIMS system for the purpose of large scale investigations of human identification mentioned in the KADAP report. LISA is easy to use, tailored made for the needs of the lab and can manage large amounts of genetic data very well. The importing tools provided by the Profile Management module are easy to use and can successfully the import/export large numbers of DNA profiles. The Kinship Analysis module is very good in building and analyzing complex pedigrees at a reasonable point. A major advantage of the Kinship Analysis module is the reporting capabilities in which analytical reports can be made showing the pedigree, the DNA profiles of all of the individuals in the pedigree, KI results, and the KI calculations for each locus [23, 31].

LISA has several disadvantages that need to be addressed because it interferes with the efficacy of using the software. One major disadvantage of LISA is its dependency on third party programs like DNA VIEW™ and Progeny©. The reason is that by using third party programs constricts FTI© to modify, improve, or correct problems within LISA because of copyright issues. FTI© would have to ask permission from the creators of these programs in order to make modifications to improve the interface between LISA and DNA VIEW™ or Progeny©. Although the copyright issue regarding these programs is only part of the problem, the real problem is the interface between the three programs which has caused several problems in LISA. Many problems regarding the Searching and Kinship Analysis modules have been traced back to

problems with the interface or within the third party programs. For example, FTI© discovered that many of the inconsistencies between KI results from LISA and DOJ VIEW 2.0 were attributed to a problem in an update for the Progeny© program. Another issue concerning the use of third party programs in LISA is that in order to set up some of the analytical settings and parameter for analysis has to be done manually in their respective programs. For example, to set up the population databases for use in kinship analysis in LISA has to be done manually in DNA VIEW™! Some of the settings and parameters in these programs cannot be set or changed by the analyst or by FTI© because the creators of these programs have sole control over these settings. The use of third party programs inhibits the efficient use of LISA making it a lot less user friendly. FTI© would benefit if they had their own designed analytical and pedigree construction software because they would have sole control of it and can make the necessary modifications, improvements, and corrections needed to make LISA an efficient LIMS system [2, 31].

Another problem is the updates and maintenance for LISA because it causes more problems than resolving them. During the course of the study, FTI© has updated LISA several times in new releases that would correct the reported software problems and improve the efficiency of LISA. In almost every occasion when LISA has been updated new problems would arise causing a week long delays in the study because LISA was either malfunctioning or inoperable. As it is installing updates for LISA can take as long 1 to 2 days. If an update caused a problem that made LISA inoperable, the problem would be immediately reported to FTI© which took them 2 to 3 days to review and

identify the problem. To correct the problem, it took FTI© less than a day to correct it. The problems with updates is important because a laboratory can potentially loose around a week in analyzing data especially in the middle of an investigation all because of a routine update that made LISA inoperable. To avoid this problem, FTI© has to figure out is the source of the problem with the updates for LISA and correct it. Another suggestion is to plan a routine maintenance schedule with FTI© in order to prevent interruptions in casework flow for a lab that uses LISA. A lab implementing LISA should develop a contingency plan consisting of protocols that use back up programs while LISA is inoperable for a brief period of time. Constant communication is emphasized between the lab and FTI to report, and resolve problems encountered in LISA in every day use.

A third major problem regarding LISA is that it has too many software problems that are hard to find or duplicate. However, finding such problems can be done easily through validation studies which can help identify such problems that can inhibit efficient use of LISA. Another problem is that LISA is a tailored made program which is also a good thing but an economically restrained lab may miss out in some useful features. A lab that is considering in implementing LISA must figure out the operational needs of the lab in order to make the right decisions in selecting what features they need in a LIMS system. Other minute problems found in LISA is that there is no advanced features for pedigree design, KI values cannot be saved in LISA, certain population databases have not been installed in the software, and no user manual has been developed. LISA as a whole has lot of potential in becoming a powerful LIMS system that can manage and analyze large quantities of genetic data despite the problems it has. It is up to FTI© to

resolve these problems in order to improve the overall efficiency of LISA as a tool for human identification of unknown human remains.

APPENDIX

APPENDIX

APPENDIX A

NUMBER OF OBSERVATIONS PER POPULATION GROUP VERSUS THE NATURAL LOGARITHMN OF THE KINSHIP INDEX FOR EACH SCENARIO

Category 1 Scenarios

Scenario 1

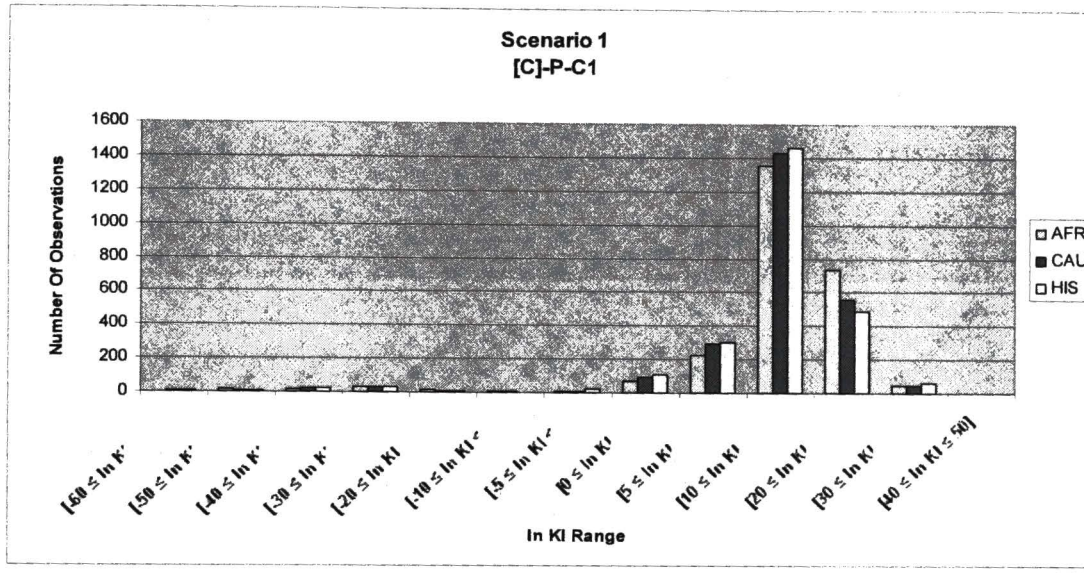


Figure 1-1: The distribution of the number observations per range of the ln KI for each population group for Scenario 1.

Scenario 1	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-60 ≤ ln KI ≤ -50]	3	5	5
[-50 ≤ ln KI ≤ -40]	13	9	9
[-40 ≤ ln KI ≤ -30]	19	25	24
[-30 ≤ ln KI ≤ -20]	28	29	29
[-20 ≤ ln KI ≤ -10]	16	11	11
[-10 ≤ ln KI ≤ -5]	4	7	3
[-5 ≤ ln KI ≤ 0]	8	10	20
[0 ≤ ln KI ≤ 5]	73	90	111
[5 ≤ ln KI ≤ 10]	222	290	298
[10 ≤ ln KI ≤ 20]	1349	1431	1458
[20 ≤ ln KI ≤ 30]	733	557	484
[30 ≤ ln KI ≤ 40]	46	50	61
[40 ≤ ln KI ≤ 50]	0	0	1

Table 1-1: The number observations per range of the ln KI for each population group for Scenario 1.

Scenario 2

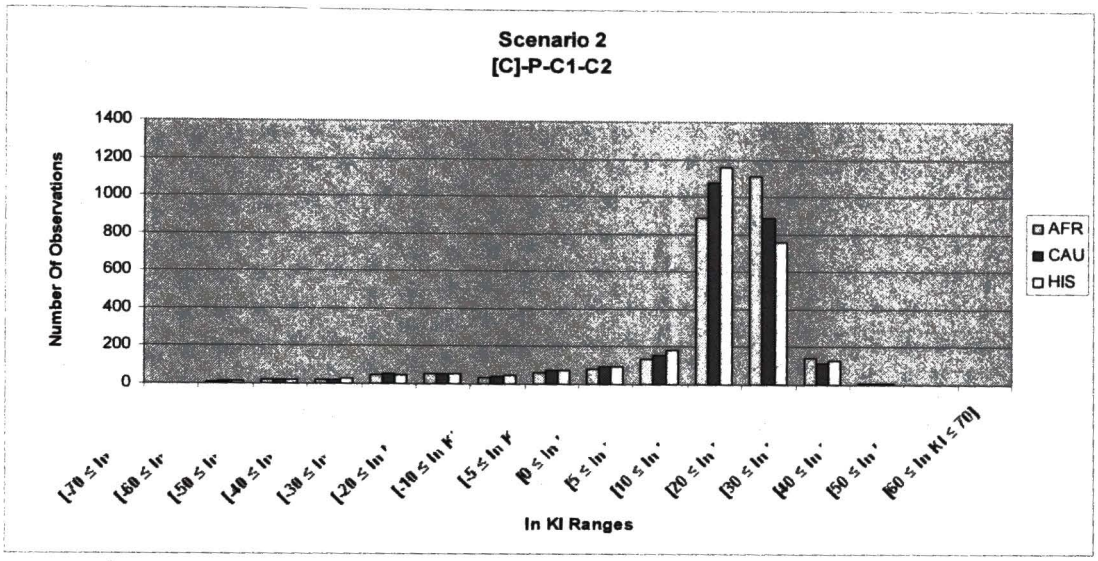


Figure 1-2: The distribution of the number observations per range of the ln KI for each population group for Scenario 2.

Scenario 2	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-70 ≤ ln KI ≤ -60]	2	1	1
[-60 ≤ ln KI ≤ -50]	8	11	11
[-50 ≤ ln KI ≤ -40]	17	18	17
[-40 ≤ ln KI ≤ -30]	22	23	25
[-30 ≤ ln KI ≤ -20]	45	51	46
[-20 ≤ ln KI ≤ -10]	55	51	56
[-10 ≤ ln KI ≤ -5]	34	41	47
[-5 ≤ ln KI ≤ 0]	61	72	73
[0 ≤ ln KI ≤ 5]	82	98	97
[5 ≤ ln KI ≤ 10]	134	156	181
[10 ≤ ln KI ≤ 20]	884	1073	1155
[20 ≤ ln KI ≤ 30]	1112	886	759
[30 ≤ ln KI ≤ 40]	139	117	128
[40 ≤ ln KI ≤ 50]	7	4	7
[50 ≤ ln KI ≤ 60]	1	1	0
[60 ≤ ln KI ≤ 70]	1	1	1

Table 1-2: The number observations per range of the ln KI for each population group for Scenario 2.

Scenario 3

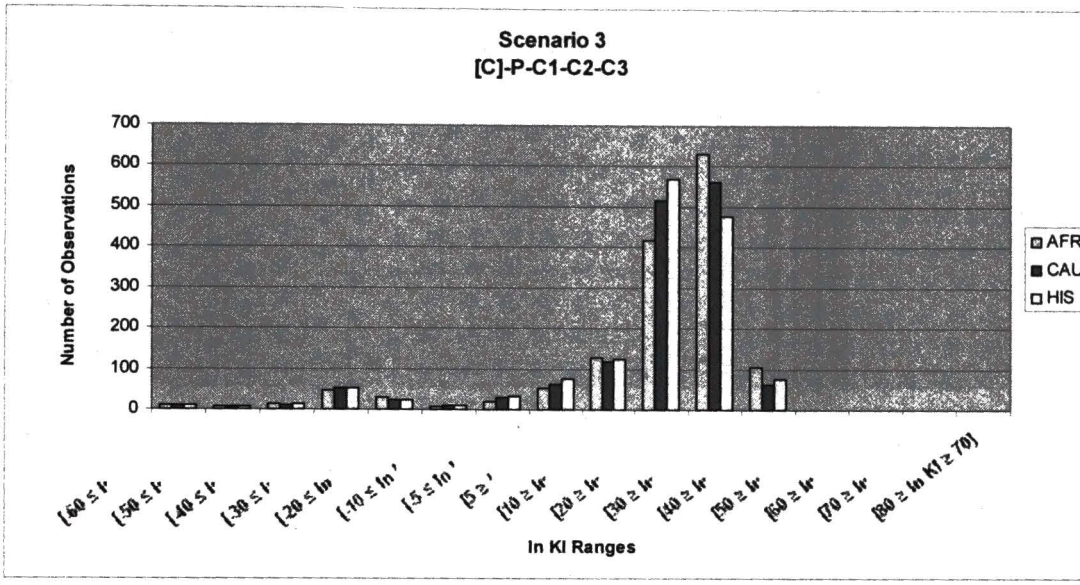


Figure 1-3: The distribution of the number observations per range of the ln KI for each population group for Scenario 3.

Scenario 3	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-60 ≤ ln KI ≤ -50]	10	10	10
[-50 ≤ ln KI ≤ -40]	6	7	7
[-40 ≤ ln KI ≤ -30]	13	11	14
[-30 ≤ ln KI ≤ -20]	45	53	51
[-20 ≤ ln KI ≤ -10]	28	22	22
[-10 ≤ ln KI ≤ -5]	8	11	9
[-5 ≤ ln KI ≤ 0]	19	29	32
[5 ≥ ln KI ≥ 0]	51	63	75
[10 ≥ ln KI ≥ 05]	128	118	126
[20 ≥ ln KI ≥ 10]	419	516	568
[30 ≥ ln KI ≥ 20]	631	561	475
[40 ≥ ln KI ≥ 30]	105	62	74
[50 ≥ ln KI ≥ 40]	0	0	0
[60 ≥ ln KI ≥ 50]	0	0	0
[70 ≥ ln KI ≥ 60]	0	0	0
[80 ≥ ln KI ≥ 70]	1	1	1

Table 1-3: The number observations per range of the ln KI for each population group for Scenario 3.

Scenario 4

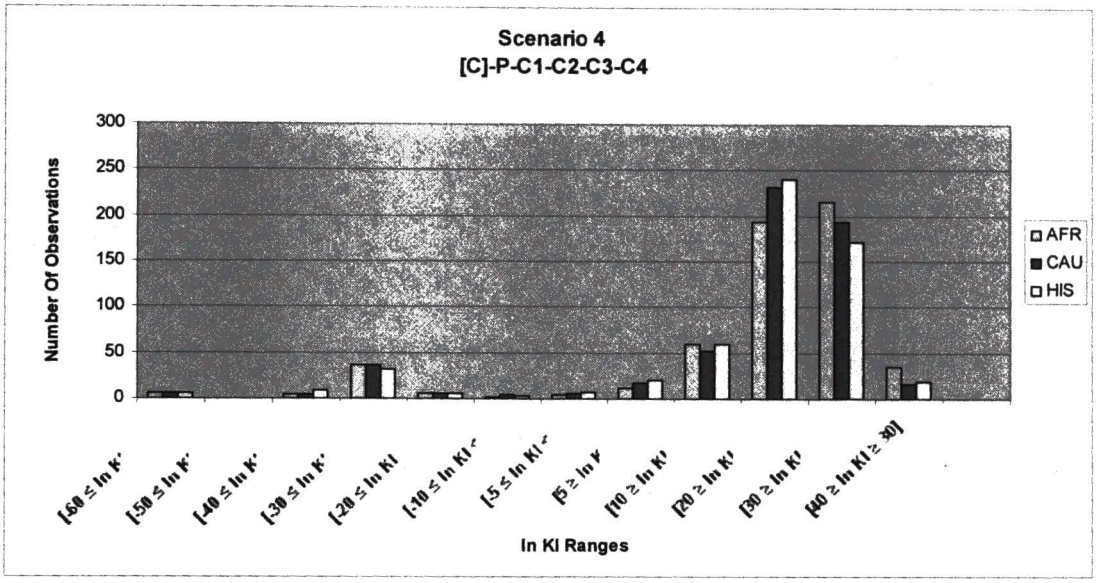


Figure 1-4: The distribution of the number observations per range of the ln KI for each population group for Scenario 4.

Scenario 4	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-60 ≤ ln KI ≤ -50]	6	6	6
[-50 ≤ ln KI ≤ -40]	0	0	0
[-40 ≤ ln KI ≤ -30]	5	5	9
[-30 ≤ ln KI ≤ -20]	36	36	32
[-20 ≤ ln KI ≤ -10]	6	6	6
[-10 ≤ ln KI ≤ -5]	1	4	3
[-5 ≤ ln KI ≤ 0]	5	6	8
[5 ≥ ln KI ≥ 0]	12	18	20
[10 ≥ ln KI ≥ 05]	60	53	60
[20 ≥ ln KI ≥ 10]	194	231	240
[30 ≥ ln KI ≥ 20]	215	194	172
[40 ≥ ln KI ≥ 30]	35	16	19

Table 1-4: The number observations per range of the ln KI for each population group for Scenario 4.

Scenario 5

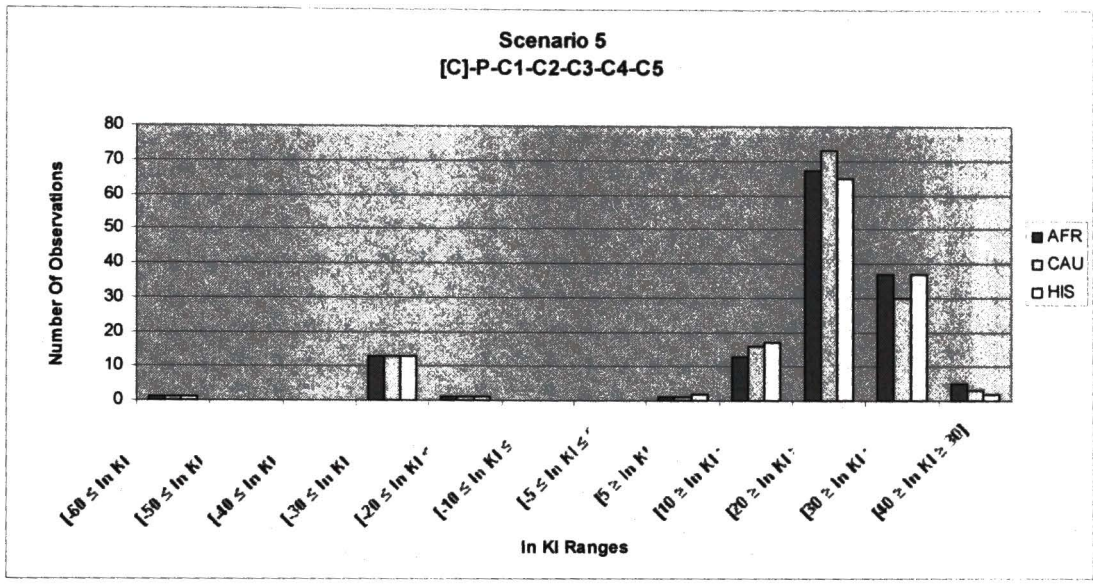


Figure 1-5: The distribution of the number observations per range of the ln KI for each population group for Scenario 5.

Scenario 5	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-60 ≤ ln KI ≤ -50]	1	1	1
[-50 ≤ ln KI ≤ -40]	0	0	0
[-40 ≤ ln KI ≤ -30]	0	0	0
[-30 ≤ ln KI ≤ -20]	13	13	13
[-20 ≤ ln KI ≤ -10]	1	1	1
[-10 ≤ ln KI ≤ -5]	0	0	0
[-5 ≤ ln KI ≤ 0]	0	0	0
[5 ≥ ln KI ≥ 0]	1	1	2
[10 ≥ ln KI ≥ 05]	13	16	17
[20 ≥ ln KI ≥ 10]	67	73	65
[30 ≥ ln KI ≥ 20]	37	30	37
[40 ≥ ln KI ≥ 30]	5	3	2

Table 1-5: The number observations per range of the ln KI for each population group for Scenario 5.

Scenario 6

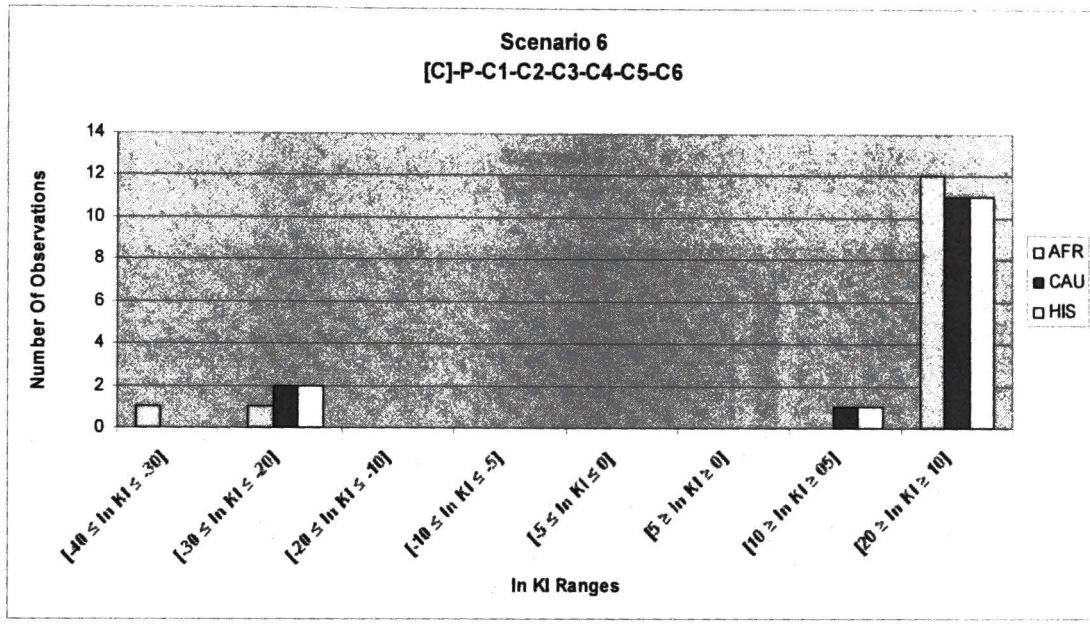


Figure 1-6: The distribution of the number observations per range of the ln KI for each population group for Scenario 6.

Scenario 6	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	1	0	0
[-30 ≤ ln KI ≤ -20]	1	2	2
[-20 ≤ ln KI ≤ -10]	0	0	0
[-10 ≤ ln KI ≤ -5]	0	0	0
[-5 ≤ ln KI ≤ 0]	0	0	0
[5 ≥ ln KI ≥ 0]	0	0	0
[10 ≥ ln KI ≥ 05]	0	1	1
[20 ≥ ln KI ≥ 10]	12	11	11

Table 1-6: The number observations per range of the ln KI for each population group for Scenario 6.

Category 2 Scenarios

Scenario 7

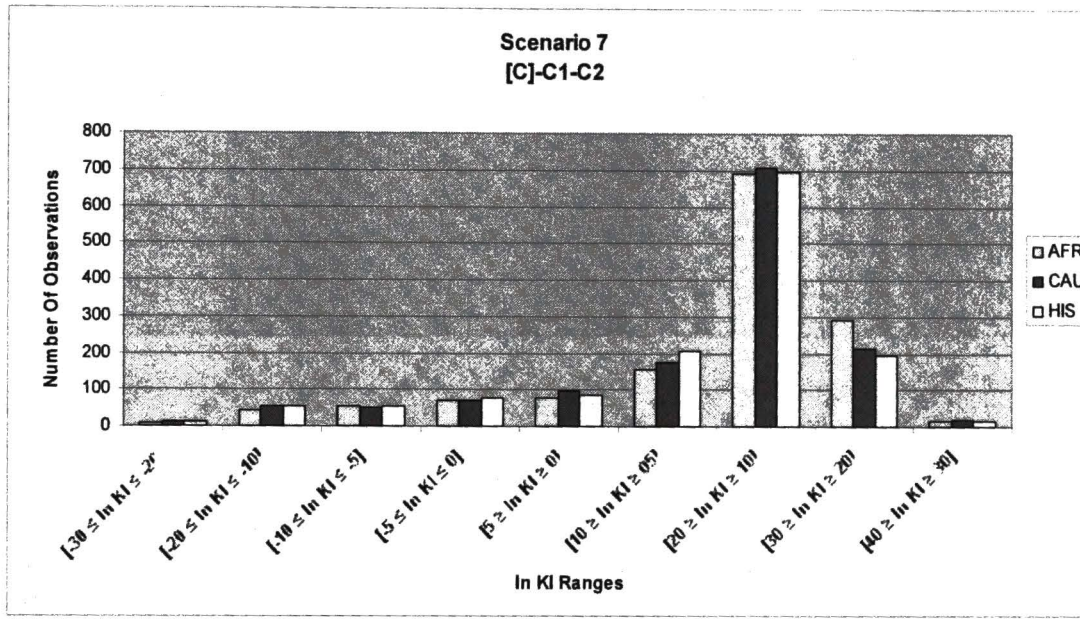


Figure 2-7: The distribution of the number observations per range of the ln KI for each population group for Scenario 7.

Scenario 7	Number Of Observations Per population Group		
ln KI Ranges	AFR	CAU	HIS
[-30 ≤ ln KI ≤ -20]	9	12	12
[-20 ≤ ln KI ≤ -10]	42	55	53
[-10 ≤ ln KI ≤ -5]	53	52	55
[-5 ≤ ln KI ≤ 0]	69	70	79
[5 ≥ ln KI ≥ 0]	78	99	87
[10 ≥ ln KI ≥ 05]	155	176	207
[20 ≥ ln KI ≥ 10]	693	707	697
[30 ≥ ln KI ≥ 20]	290	214	196
[40 ≥ ln KI ≥ 30]	14	18	17

Table 2-7: The number observations per range of the ln KI for each population group for Scenario 7.

Scenario 8

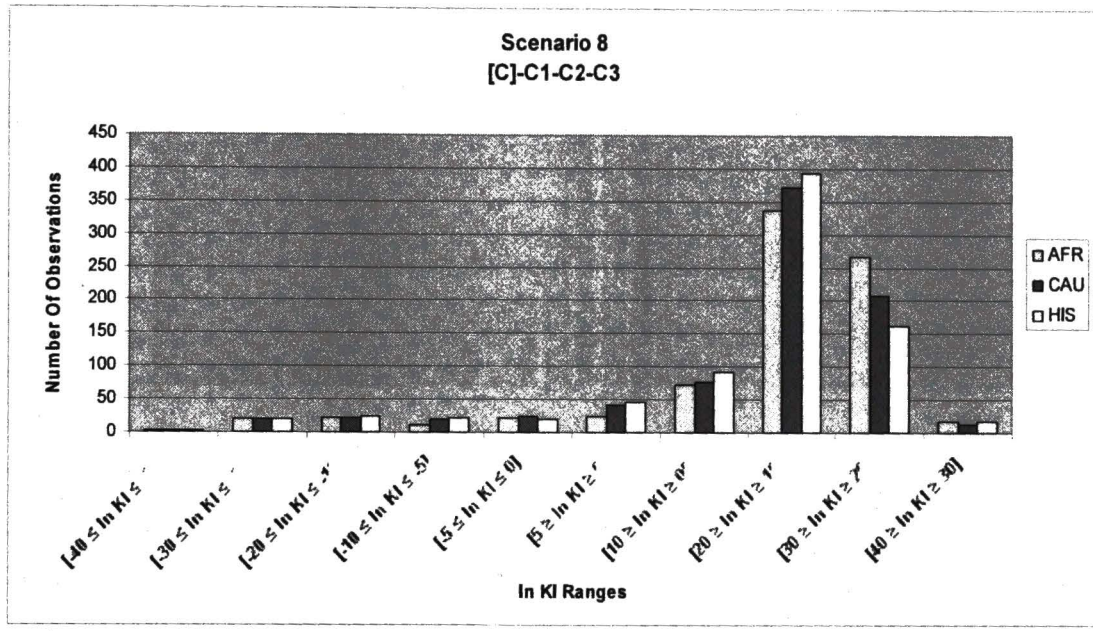


Figure 2-8: The distribution of the number observations per range of the ln KI for each population group for Scenario 8.

Scenario 8	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	3	2	2
[-30 ≤ ln KI ≤ -20]	20	19	20
[-20 ≤ ln KI ≤ -10]	22	22	24
[-10 ≤ ln KI ≤ -5]	12	19	21
[-5 ≤ ln KI ≤ 0]	22	24	19
[5 ≥ ln KI ≥ 0]	24	42	46
[10 ≥ ln KI ≥ 05]	72	76	91
[20 ≥ ln KI ≥ 10]	336	372	394
[30 ≥ ln KI ≥ 20]	267	207	161
[40 ≥ ln KI ≥ 30]	18	13	18

Table 2-8: The number observations per range of the ln KI for each population group for Scenario 8.

Scenario 9

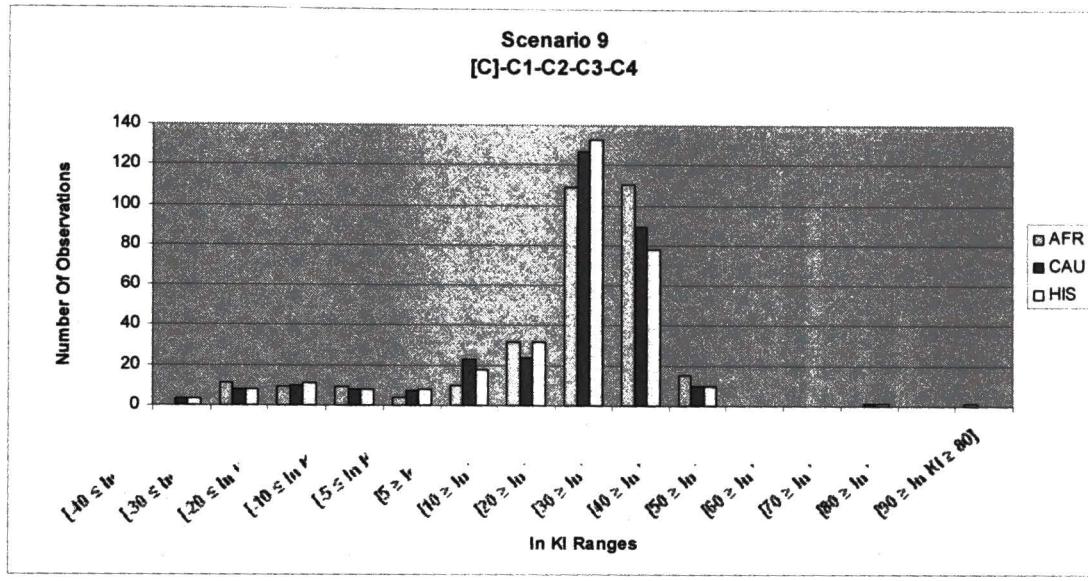


Figure 2-9: The distribution of the number observations per range of the ln KI for each population group for Scenario 9.

Scenario 9	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	0	3	3
[-30 ≤ ln KI ≤ -20]	11	8	8
[-20 ≤ ln KI ≤ -10]	9	10	11
[-10 ≤ ln KI ≤ -5]	9	8	8
[-5 ≤ ln KI ≤ 0]	4	7	8
[5 ≥ ln KI ≥ 0]	10	23	18
[10 ≥ ln KI ≥ 05]	32	24	32
[20 ≥ ln KI ≥ 10]	109	127	133
[30 ≥ ln KI ≥ 20]	110	89	78
[40 ≥ ln KI ≥ 30]	15	10	10
[50 ≥ ln KI ≥ 40]	0	0	0
[60 ≥ ln KI ≥ 50]	0	0	0
[70 ≥ ln KI ≥ 60]	0	1	1
[80 ≥ ln KI ≥ 70]	0	0	0
[90 ≥ ln KI ≥ 80]	1	0	0

Table 2-9: The number observations per range of the ln KI for each population group for Scenario 9.

Scenario 10

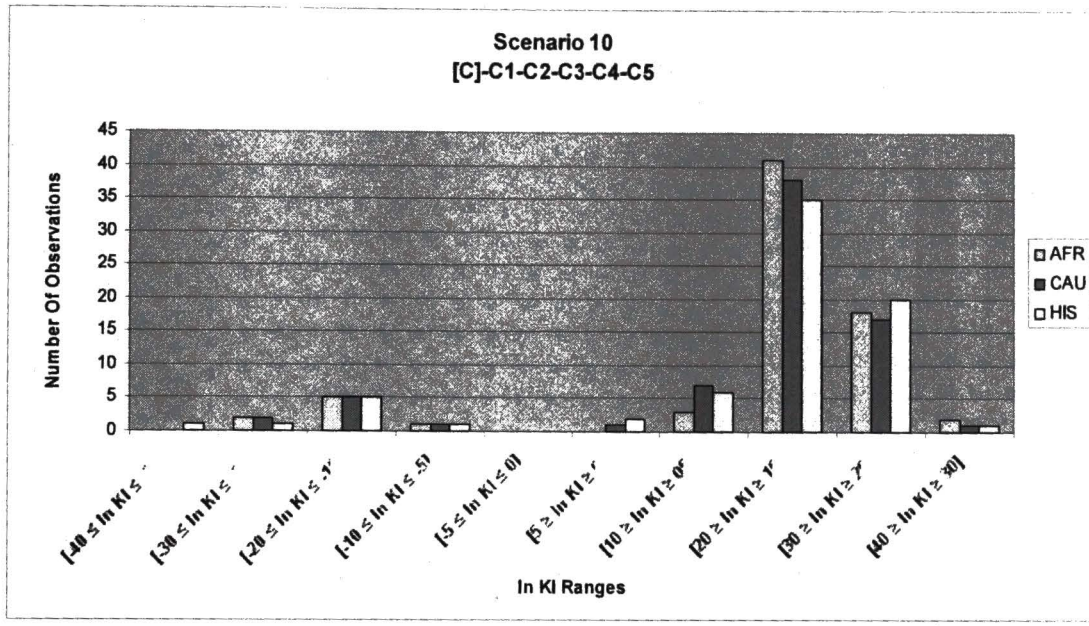


Figure 2-10: The distribution of the number observations per range of the ln KI for each population group for Scenario 10.

Scenario 10	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	0	0	1
[-30 ≤ ln KI ≤ -20]	2	2	1
[-20 ≤ ln KI ≤ -10]	5	5	5
[-10 ≤ ln KI ≤ -5]	1	1	1
[-5 ≤ ln KI ≤ 0]	0	0	0
[5 ≥ ln KI ≥ 0]	0	1	2
[10 ≥ ln KI ≥ 05]	3	7	6
[20 ≥ ln KI ≥ 10]	41	38	35
[30 ≥ ln KI ≥ 20]	18	17	20
[40 ≥ ln KI ≥ 30]	2	1	1

Table 2-10: The number observations per range of the ln KI for each population group for Scenario 10.

Scenario 11

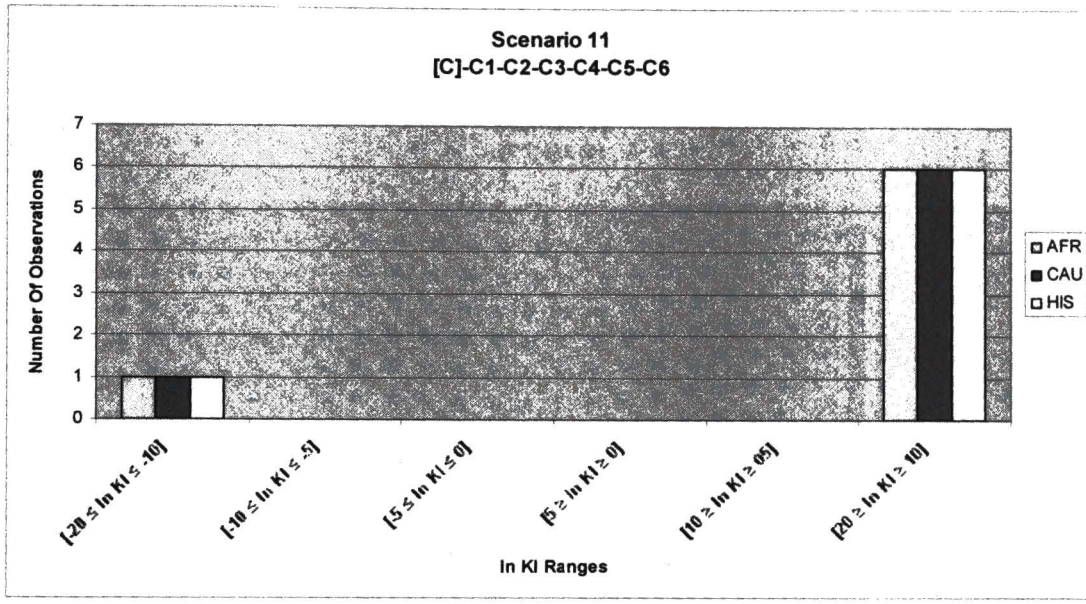


Figure 2-11: The distribution of the number observations per range of the ln KI for each population group for Scenario 11.

Scenario 11	Number Of Observations Per population Group		
ln KI Ranges	AFR	CAU	HIS
[-20 ≤ ln KI ≤ -10]	1	1	1
[-10 ≤ ln KI ≤ -5]	0	0	0
[-5 ≤ ln KI ≤ 0]	0	0	0
[5 ≥ ln KI ≥ 0]	0	0	0
[10 ≥ ln KI ≥ 05]	0	0	0
[20 ≥ ln KI ≥ 10]	6	6	6

Table 2-11: The number observations per range of the ln KI for each population group for Scenario 11.

Category 3 Scenarios

Scenario 12

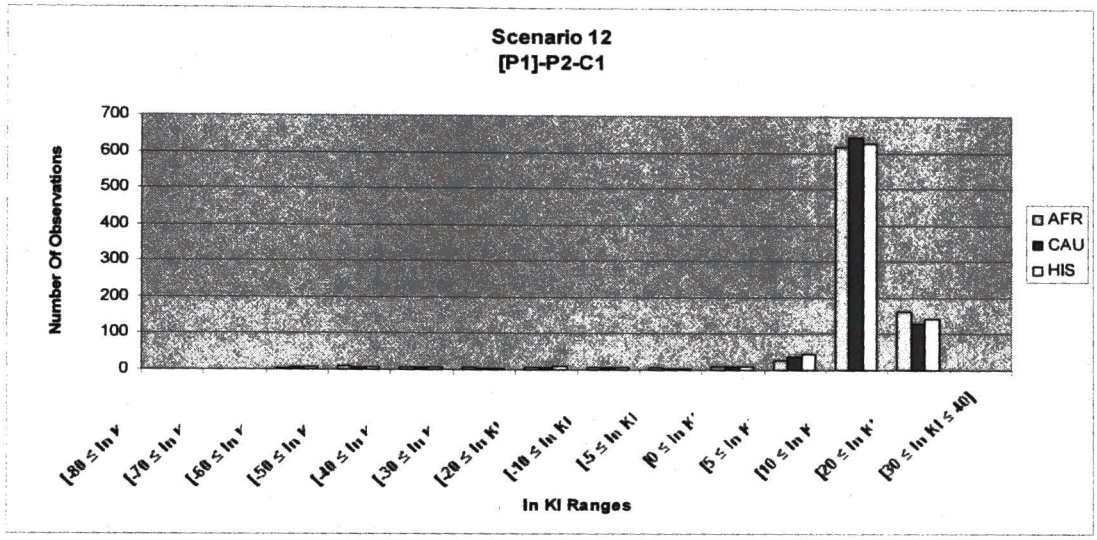


Figure 3-12: The distribution of the number observations per range of the ln KI for each population group for Scenario 12.

Scenario 12	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-80 ≤ ln KI ≤ -70]	1	1	1
[-70 ≤ ln KI ≤ -60]	1	1	1
[-60 ≤ ln KI ≤ -50]	5	6	6
[-50 ≤ ln KI ≤ -40]	9	8	8
[-40 ≤ ln KI ≤ -30]	6	6	6
[-30 ≤ ln KI ≤ -20]	6	5	4
[-20 ≤ ln KI ≤ -10]	6	8	9
[-10 ≤ ln KI ≤ -5]	7	7	7
[-5 ≤ ln KI ≤ 0]	7	4	4
[5 ≥ ln KI ≥ 0]	9	10	9
[10 ≥ ln KI ≥ 05]	26	37	43
[20 ≥ ln KI ≥ 10]	617	642	624
[30 ≥ ln KI ≥ 20]	164	128	141
[40 ≥ ln KI ≥ 30]	0	1	1

Table 3-12: The number observations per range of the ln KI for each population group for Scenario 12.

Scenario 13

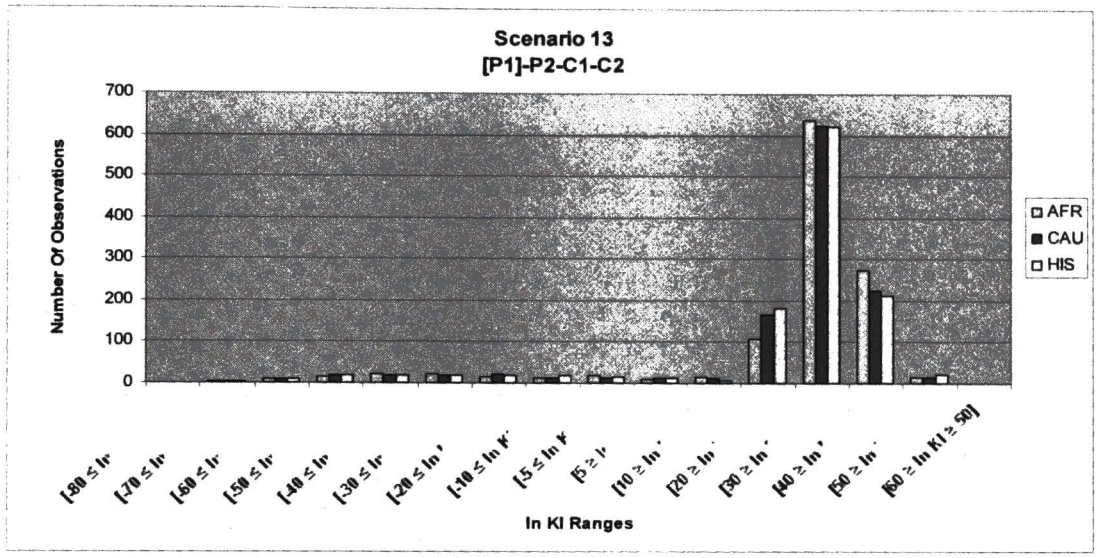


Figure 3-13: The distribution of the number observations per range of the ln KI for each population group for Scenario 13.

Scenario 13	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-80 ≤ ln KI ≤ -70]	1	1	1
[-70 ≤ ln KI ≤ -60]	3	3	3
[-60 ≤ ln KI ≤ -50]	10	10	10
[-50 ≤ ln KI ≤ -40]	15	18	18
[-40 ≤ ln KI ≤ -30]	20	19	18
[-30 ≤ ln KI ≤ -20]	21	18	19
[-20 ≤ ln KI ≤ -10]	15	21	17
[-10 ≤ ln KI ≤ -5]	12	13	17
[-5 ≤ ln KI ≤ 0]	17	13	15
[5 ≥ ln KI ≥ 0]	8	11	13
[10 ≥ ln KI ≥ 05]	16	13	7
[20 ≥ ln KI ≥ 10]	108	167	180
[30 ≥ ln KI ≥ 20]	634	622	619
[40 ≥ ln KI ≥ 30]	274	224	212
[50 ≥ ln KI ≥ 40]	16	16	20
[60 ≥ ln KI ≥ 50]	0	1	1

Table 3-13: The number observations per range of the ln KI for each population group for Scenario 13.

Scenario 14

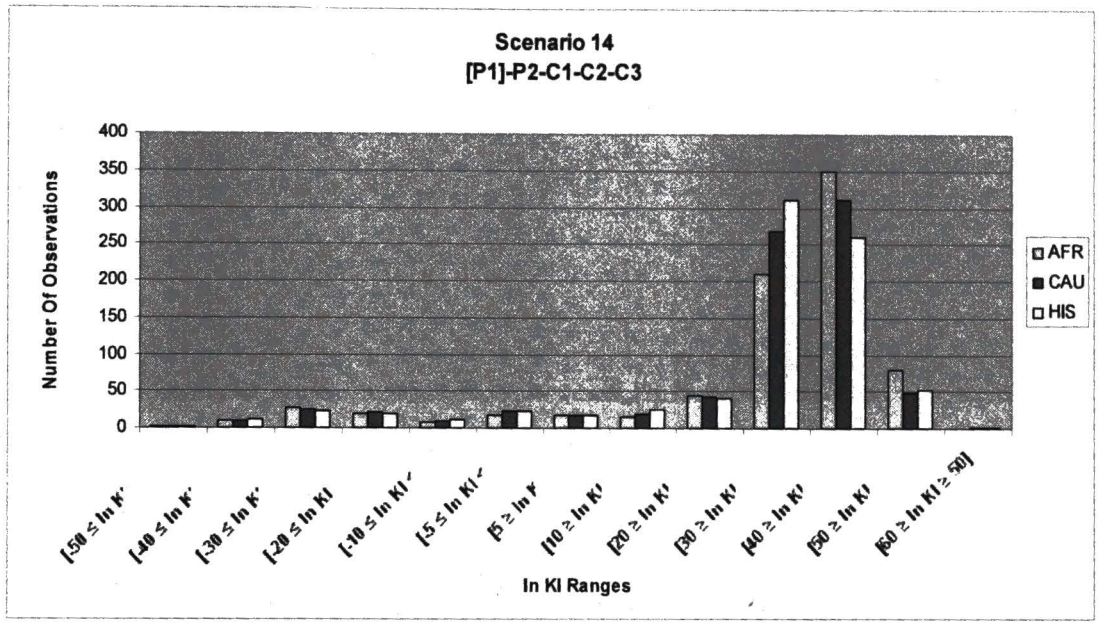


Figure 3-14: The distribution of the number observations per range of the ln KI for each population group for Scenario 14.

Scenario 14	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-50 ≤ ln KI ≤ -40]	2	2	2
[-40 ≤ ln KI ≤ -30]	9	10	12
[-30 ≤ ln KI ≤ -20]	28	26	24
[-20 ≤ ln KI ≤ -10]	20	21	19
[-10 ≤ ln KI ≤ -5]	7	9	11
[-5 ≤ ln KI ≤ 0]	17	23	23
[5 ≥ ln KI ≥ 0]	18	18	17
[10 ≥ ln KI ≥ 05]	15	19	25
[20 ≥ ln KI ≥ 10]	44	42	41
[30 ≥ ln KI ≥ 20]	209	267	310
[40 ≥ ln KI ≥ 30]	349	311	261
[50 ≥ ln KI ≥ 40]	80	49	52
[60 ≥ ln KI ≥ 50]	0	1	1

Table 3-14: The number observations per range of the ln KI for each population group for Scenario 14.

Scenario 15

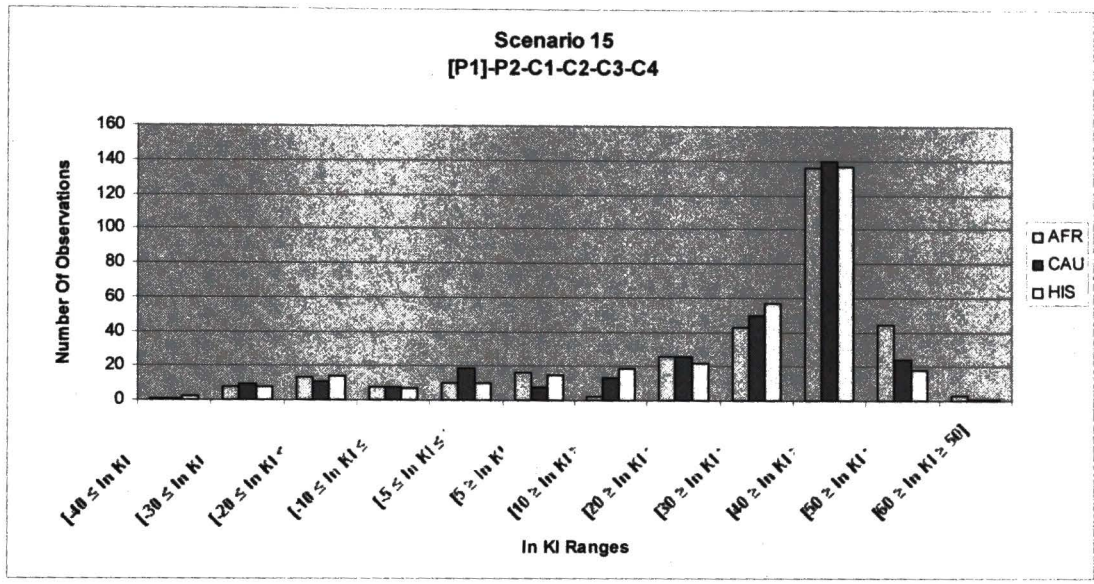


Figure 3-15: The distribution of the number observations per range of the ln KI for each population group for Scenario 15.

Scenario 15	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	1	1	2
[-30 ≤ ln KI ≤ -20]	8	9	8
[-20 ≤ ln KI ≤ -10]	13	11	14
[-10 ≤ ln KI ≤ -5]	8	8	7
[-5 ≤ ln KI ≤ 0]	10	19	10
[5 ≥ ln KI ≥ 0]	16	8	15
[10 ≥ ln KI ≥ 05]	2	13	19
[20 ≥ ln KI ≥ 10]	26	26	22
[30 ≥ ln KI ≥ 20]	43	50	57
[40 ≥ ln KI ≥ 30]	136	140	137
[50 ≥ ln KI ≥ 40]	44	24	18
[60 ≥ ln KI ≥ 50]	3	1	1

Table 3-15: The number observations per range of the ln KI for each population group for Scenario 15.

Scenario 16

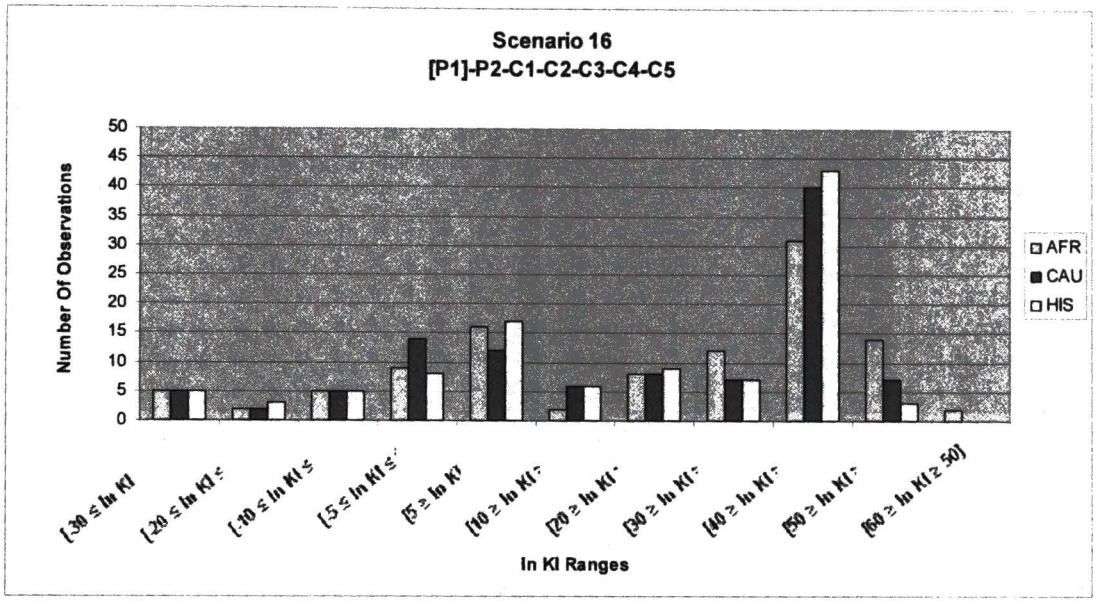


Figure 3-16: The distribution of the number observations per range of the ln KI for each population group for Scenario 16.

Scenario 16	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-30 ≤ ln KI ≤ -20]	5	5	5
[-20 ≤ ln KI ≤ -10]	2	2	3
[-10 ≤ ln KI ≤ -5]	5	5	5
[-5 ≤ ln KI ≤ 0]	9	14	8
[5 ≥ ln KI ≥ 0]	16	12	17
[10 ≥ ln KI ≥ 05]	2	6	6
[20 ≥ ln KI ≥ 10]	8	8	9
[30 ≥ ln KI ≥ 20]	12	7	7
[40 ≥ ln KI ≥ 30]	31	40	43
[50 ≥ ln KI ≥ 40]	14	7	3
[60 ≥ ln KI ≥ 50]	2	0	0

Table 3-16: The number observations per range of the ln KI for each population group for Scenario 16.

Scenario 17

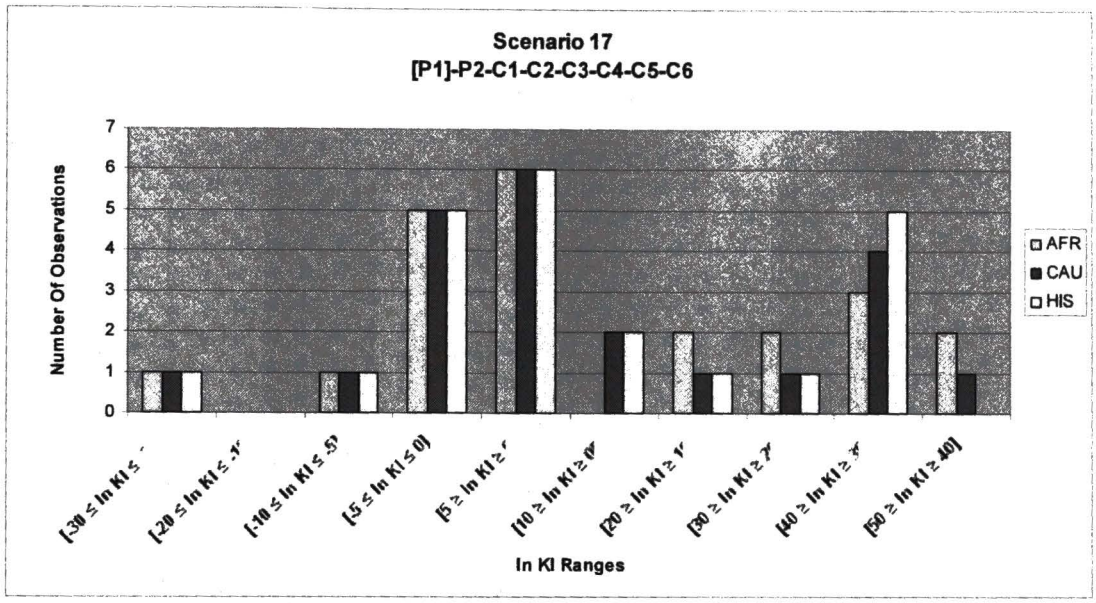


Figure 3-17: The distribution of the number observations per range of the ln KI for each population group for Scenario 17.

Scenario 17	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-30 ≤ ln KI ≤ -20]	1	1	1
[-20 ≤ ln KI ≤ -10]	0	0	0
[-10 ≤ ln KI ≤ -5]	1	1	1
[-5 ≤ ln KI ≤ 0]	5	5	5
[5 ≥ ln KI ≥ 0]	6	6	6
[10 ≥ ln KI ≥ 05]	0	2	2
[20 ≥ ln KI ≥ 10]	2	1	1
[30 ≥ ln KI ≥ 20]	2	1	1
[40 ≥ ln KI ≥ 30]	3	4	5
[50 ≥ ln KI ≥ 40]	2	1	0

Table 3-17: The number observations per range of the ln KI for each population group for Scenario 17.

Scenario 18

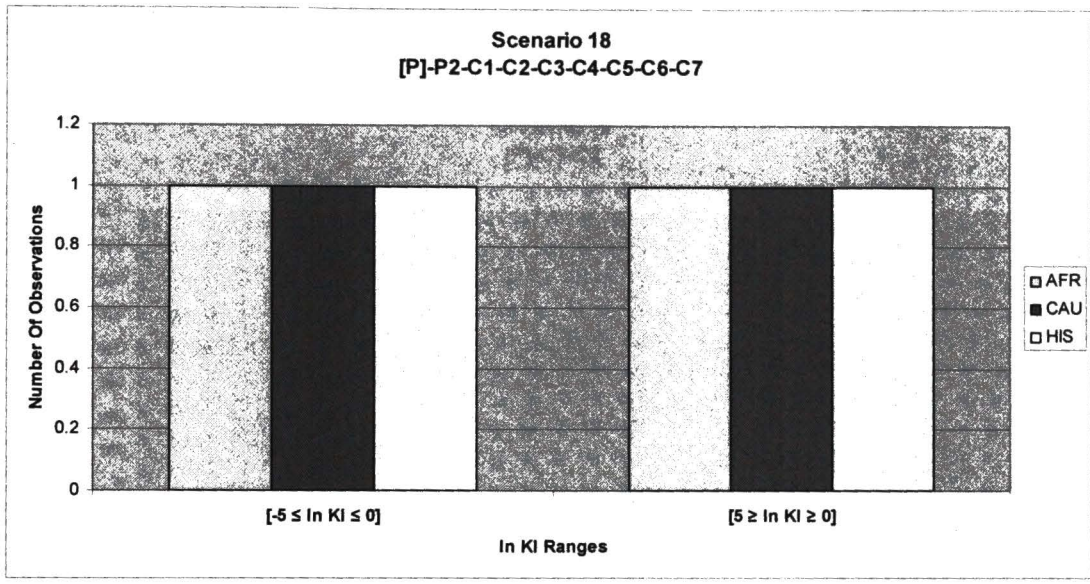


Figure 3-18: The distribution of the number observations per range of the ln KI for each population group for Scenario 18.

Scenario 18	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-5 ≤ ln KI ≤ 0]	1	1	1
[5 ≥ ln KI ≥ 0]	1	1	1

Table 3-18: The number observations per range of the ln KI for each population group for Scenario 18.

Category 4 Scenarios

Scenario 19

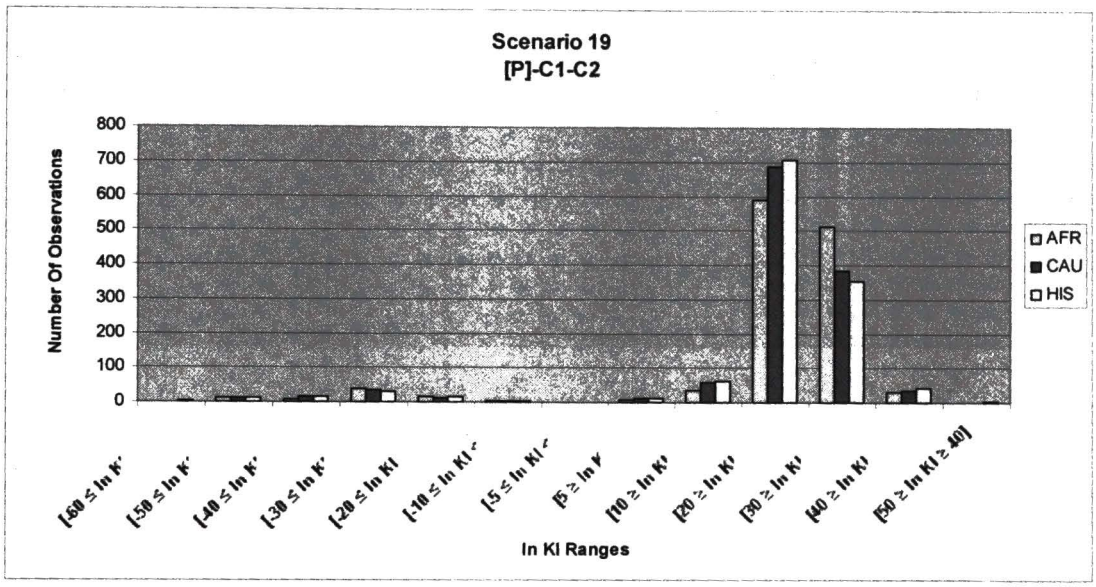


Figure 4-19: The distribution of the number observations per range of the ln KI for each population group for Scenario 19.

Scenario 19	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-60 ≤ ln KI ≤ -50]	1	1	2
[-50 ≤ ln KI ≤ -40]	13	12	12
[-40 ≤ ln KI ≤ -30]	9	15	15
[-30 ≤ ln KI ≤ -20]	38	34	31
[-20 ≤ ln KI ≤ -10]	15	12	15
[-10 ≤ ln KI ≤ -5]	3	4	3
[-5 ≤ ln KI ≤ 0]	1	1	1
[5 ≥ ln KI ≥ 0]	9	12	13
[10 ≥ ln KI ≥ 05]	34	57	61
[20 ≥ ln KI ≥ 10]	590	689	706
[30 ≥ ln KI ≥ 20]	512	385	352
[40 ≥ ln KI ≥ 30]	32	34	43
[50 ≥ ln KI ≥ 40]	0	1	3

Table 4-19: The number observations per range of the ln KI for each population group for Scenario 19.

Scenario 20

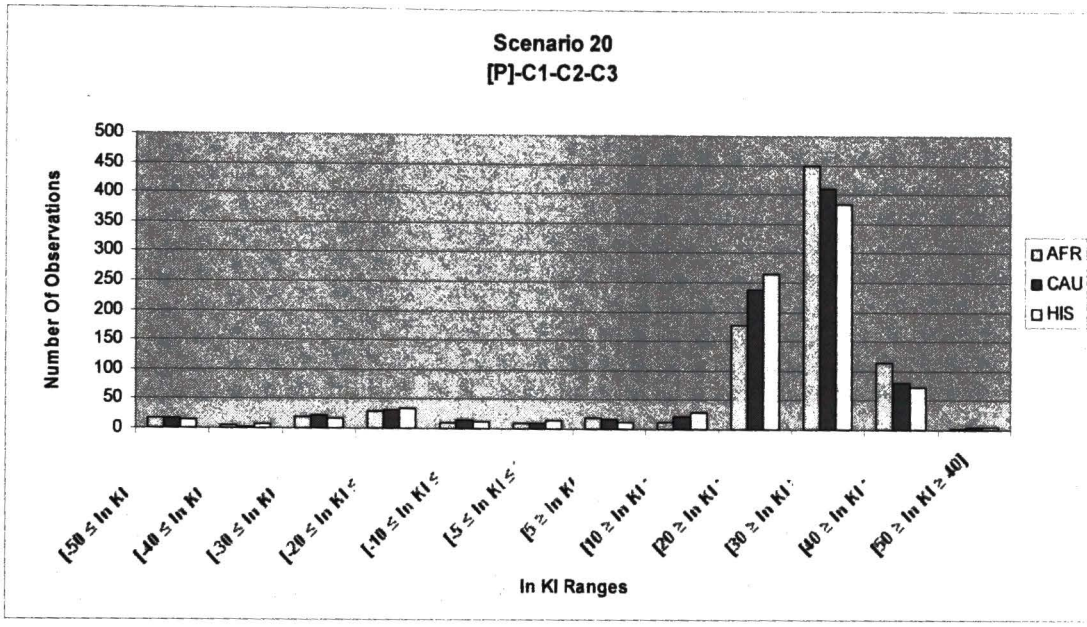


Figure 4-20: The distribution of the number observations per range of the ln KI for each population group for Scenario 20.

Scenario 20	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-50 ≤ ln KI ≤ -40]	16	16	15
[-40 ≤ ln KI ≤ -30]	6	3	7
[-30 ≤ ln KI ≤ -20]	20	22	17
[-20 ≤ ln KI ≤ -10]	30	31	34
[-10 ≤ ln KI ≤ -5]	10	14	13
[-5 ≤ ln KI ≤ 0]	10	10	14
[5 ≥ ln KI ≥ 0]	19	16	11
[10 ≥ ln KI ≥ 05]	12	22	29
[20 ≥ ln KI ≥ 10]	177	237	264
[30 ≥ ln KI ≥ 20]	450	409	384
[40 ≥ ln KI ≥ 30]	115	81	74
[50 ≥ ln KI ≥ 40]	2	6	5

Table 4-20: The number observations per range of the ln KI for each population group for Scenario 20.

Scenario 21

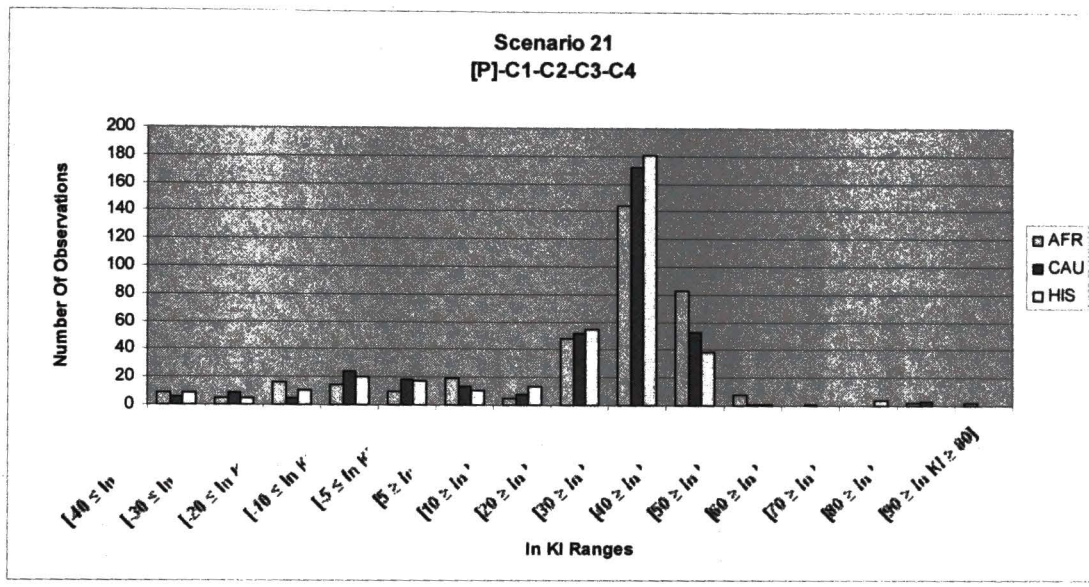


Figure 4-21: The distribution of the number observations per range of the ln KI for each population group for Scenario 21.

Scenario 21	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	9	6	9
[-30 ≤ ln KI ≤ -20]	5	9	5
[-20 ≤ ln KI ≤ -10]	16	5	11
[-10 ≤ ln KI ≤ -5]	14	24	20
[-5 ≤ ln KI ≤ 0]	10	18	17
[5 ≥ ln KI ≥ 0]	19	13	11
[10 ≥ ln KI ≥ 05]	5	8	13
[20 ≥ ln KI ≥ 10]	48	52	55
[30 ≥ ln KI ≥ 20]	144	172	181
[40 ≥ ln KI ≥ 30]	83	53	38
[50 ≥ ln KI ≥ 40]	8	1	1
[60 ≥ ln KI ≥ 50]	0	1	0
[70 ≥ ln KI ≥ 60]	0	0	4
[80 ≥ ln KI ≥ 70]	2	3	0
[90 ≥ ln KI ≥ 80]	2	0	0

Table 4-21: The number observations per range of the ln KI for each population group for Scenario 21.

Scenario 22

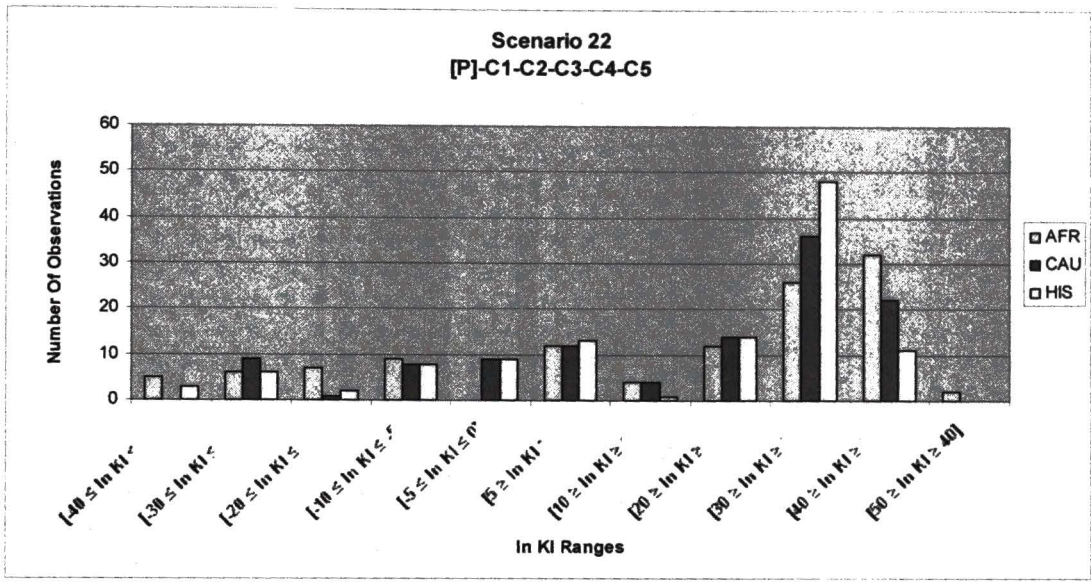


Figure 4-22: The distribution of the number observations per range of the ln KI for each population group for Scenario 22.

Scenario 22	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-40 ≤ ln KI ≤ -30]	5	0	3
[-30 ≤ ln KI ≤ -20]	6	9	6
[-20 ≤ ln KI ≤ -10]	7	1	2
[-10 ≤ ln KI ≤ -5]	9	8	8
[-5 ≤ ln KI ≤ 0]	0	9	9
[5 ≥ ln KI ≥ 0]	12	12	13
[10 ≥ ln KI ≥ 05]	4	4	1
[20 ≥ ln KI ≥ 10]	12	14	14
[30 ≥ ln KI ≥ 20]	26	36	48
[40 ≥ ln KI ≥ 30]	32	22	11
[50 ≥ ln KI ≥ 40]	2	0	0

Table 4-22: The number observations per range of the ln KI for each population group for Scenario 22.

Scenario 23

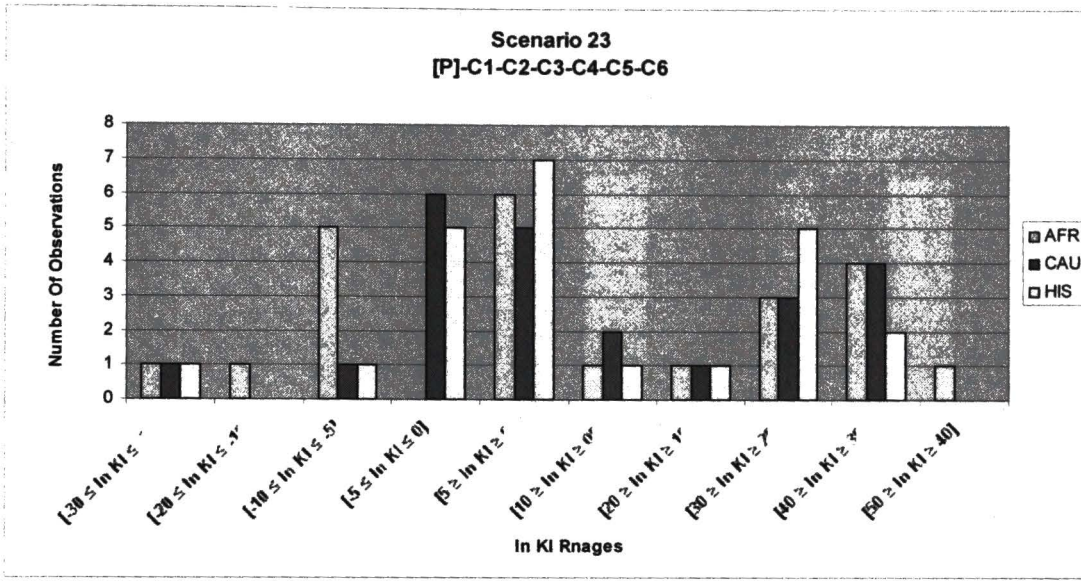


Figure 4-23: The distribution of the number observations per range of the ln KI for each population group for Scenario 23.

Scenario 23	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-30 ≤ ln KI ≤ -20]	1	1	1
[-20 ≤ ln KI ≤ -10]	1	0	0
[-10 ≤ ln KI ≤ -5]	5	1	1
[-5 ≤ ln KI ≤ 0]	0	6	5
[5 ≥ ln KI ≥ 0]	6	5	7
[10 ≥ ln KI ≥ 05]	1	2	1
[20 ≥ ln KI ≥ 10]	1	1	1
[30 ≥ ln KI ≥ 20]	3	3	5
[40 ≥ ln KI ≥ 30]	4	4	2
[50 ≥ ln KI ≥ 40]	1	0	0

Table 4-23: The number observations per range of the ln KI for each population group for Scenario 23.

Scenario 24

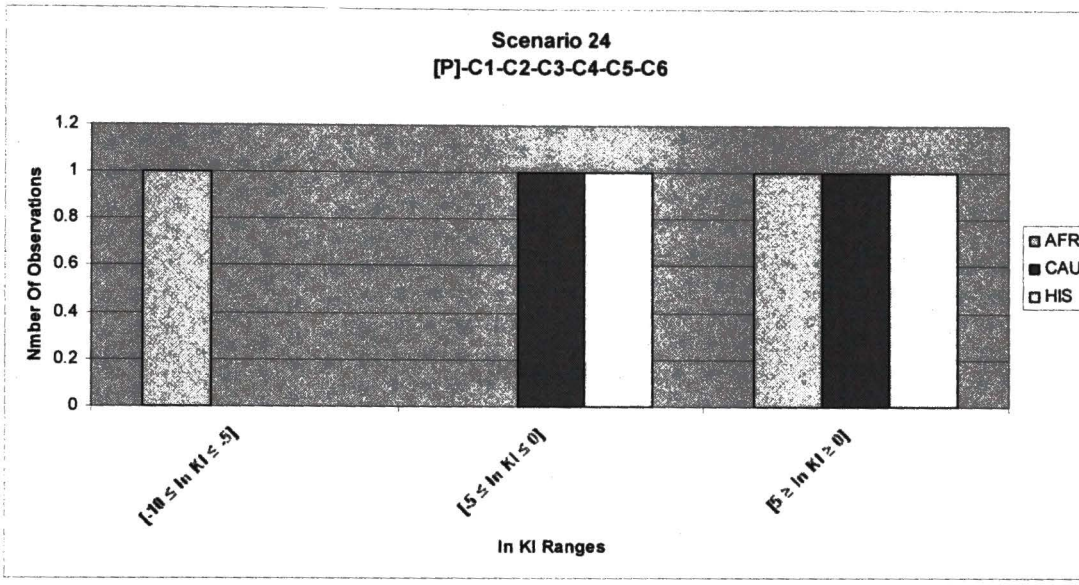


Figure 4-24: The distribution of the number observations per range of the ln KI for each population group for Scenario 24.

Scenario 24	Number Of Observations Per Population Group		
ln KI Ranges	AFR	CAU	HIS
[-10 ≤ ln KI ≤ -5]	1	0	0
[-5 ≤ ln KI ≤ 0]	0	1	1
[5 ≥ ln KI ≥ 0]	1	1	1

Table 4-24: The number observations per range of the ln KI for each population group for Scenario 24.

APPENDIX B

KI DATA FOR FAMILIES
00-0004 AND 00-0033

			Full Sib (00-0004)		Half Sib (00-0033)	
			HIS		HIS	
			C1-C4		C1*-C4	
Category 1						
	Missing/Test	Ref Combo	KI	ln KI	KI	ln KI
Scenario 1	C1	M-C2	3.00E+06	14.9141228	7.50E+01	4.3174881
[C]-P-C1		M-C3	1.40E+06	14.1519828	1.50E+04	9.6158055
		M-C4	6.40E+05	13.3692235	5.30E+03	8.5754621
		AF-C2	1.30E+08	18.683045	3.00E-09	-19.624654
		AF-C3	1.20E+08	18.6030023	8.70E-09	-18.559943
		AF-C4	1.00E+06	13.8155106	9.60E-10	-20.764088
	C2	M-C1	8.40E+06	15.9437423	1	0
		M-C3	6.80E+06	15.7324332	1.30E+04	9.4727046
		M-C4	4.60E+05	13.0389818	7.90E+04	11.277203
		AF-C1	1.50E+08	18.8261459	2.00E-01	-1.6094379
		AF-C3	1.10E+08	18.5159909	5.00E+05	13.122363
		AF-C4	1.20E+06	13.9978321	2.70E+03	7.9010071
	C3	M-C1	2.10E+07	16.860033	1.20E+03	7.0900768
		M-C2	3.70E+07	17.4264285	7.60E+04	11.238489
		M-C4	7.10E+04	11.1704352	2.60E+06	14.771022
		AF-C1	5.60E+07	17.8408622	7.50E+00	2.014903
		AF-C2	4.40E+07	17.5997002	6.60E+06	15.70258
		AF-C4	1.90E+07	16.7599495	7.80E+02	6.6592939
	C4	M-C1	3.30E+06	15.009433	1.30E+02	4.8675345
		M-C2	8.40E+05	13.6411572	1.40E+05	11.849398
		M-C3	2.40E+04	10.0858091	7.70E+05	13.554146
		AF-C1	9.60E+04	11.4721035	6.30E+03	8.7483049
		AF-C2	1.00E+05	11.5129255	1.40E+08	18.757153
		AF-C3	4.00E+06	15.2018049	3.10E+06	14.946913
			KI	ln KI	KI	ln KI
Scenario 2	C1	M-C2-C3	8.30E+07	18.2343512	8.80E+00	2.1747517
[C]-P-C1-C2		M-C2-C4	9.40E+07	18.3588053	1.50E-02	-4.1997051
		M-C3-C4	1.30E+08	18.683045	2.80E+02	5.6347896
		AF-C2-C3	5.90E+08	20.1956331	3.90E-08	-17.059704
		AF-C2-C4	2.10E+08	19.1626181	4.60E-09	-19.19721
		AF-C3-C4	1.00E+08	18.4206807	7.30E-08	-16.432806

	C2	M-C1-C3	4.10E+08	19.8316677	4.00E+00	1.3862944
		M-C1-C4	6.70E+07	18.0202032	1.10E-01	-2.2072749
		M-C3-C4	1.80E+08	19.0084674	1.60E+06	14.285514
		AF-C1-C3	5.20E+08	20.0693394	6.50E+02	6.4769724
		AF-C1-C4	2.60E+08	19.3761922	2.30E+00	0.8329091
		AF-C3-C4	9.10E+07	18.3263701	1.70E+07	16.648724
	C3	M-C1-C2	1.00E+09	20.7232658	9.00E+03	9.1049799
		M-C1-C4	1.50E+07	16.5235608	1.40E+05	11.849398
		M-C2-C4	2.70E+07	17.1113474	5.20E+07	17.766754
		AF-C1-C2	1.90E+08	19.0625346	2.40E+04	10.085809
		AF-C1-C4	2.00E+09	21.416413	1.00E+01	2.3025851
		AF-C2-C4	1.40E+09	21.0597381	4.80E+06	15.384126
	C4	M-C1-C2	2.60E+07	17.0736071	9.10E+01	4.5108595
		M-C1-C3	2.30E+06	14.6484197	1.50E+04	9.6158055
		M-C2-C3	6.20E+05	13.3374748	7.70E+05	13.554146
		AF-C1-C2	1.60E+05	11.9829291	1.20E+05	11.695247
		AF-C1-C3	2.30E+06	14.6484197	1.40E+04	9.5468126
		AF-C2-C3	3.30E+06	15.009433	1.00E+08	18.420681
			KI	ln KI	KI	ln KI
Scenario 3 [C]-P-C1-C2-C3	C1	M-C2-C3-C4	7.10E+09	22.6833606	2.90E-01	-1.2378744
		AF-C2-C3-C4	3.20E+08	19.5838316	5.40E-08	-16.734282
	C2	M-C1-C3-C4	9.40E+09	22.9639755	4.30E+01	3.7612001
		AF-C1-C3-C4	2.80E+08	19.4503002	2.30E+03	7.7406644
	C3	M-C1-C2-C4	2.10E+09	21.4652032	6.70E+06	15.717618
		AF-C1-C2-C4	2.10E+09	21.4652032	1.10E+04	9.3056506
	C4	M-C1-C2-C3	5.30E+07	17.7858025	3.80E+05	12.847927
		AF-C1-C2-C3	1.80E+06	14.4032972	3.80E+05	12.847927
Category 2						
			KI	ln KI	KI	ln KI
Scenario 7 [C]-C1-C2	C1	C2-C3	6.50E+07	17.9898978	7.20E+01	4.2766661
		C2-C4	1.80E+07	16.7058823	3.50E-02	-3.3524072
		C3-C4	1.30E+07	16.3804599	8.20E+00	2.1041342
	C2	C1-C3	4.50E+07	17.622173	3.50E+03	8.1605182
		C1-C4	2.30E+04	10.0432495	6.80E+00	1.9169226
		C3-C4	5.70E+05	13.2533916	1.00E+06	13.815511

	C3	C1-C2	1.20E+07	16.3004172	6.10E+05	13.321214
		C1-C4	1.50E+06	14.2209757	2.90E+03	7.972466
		C2-C4	3.40E+06	15.039286	1.90E+06	14.457364
	C4	C1-C2	2.40E+04	10.0858091	3.50E+02	5.8579332
		C1-C3	4.10E+05	12.9239124	8.80E+02	6.7799219
		C2-C3	2.50E+04	10.1266311	2.20E+06	14.603968
			KI	ln KI	KI	ln KI
Scenario 8 [C]-C1-C2-C3	C1	C2-C3-C4	9.40E+08	20.6613904	1.20E+00	0.1823216
	C2	C1-C3-C4	3.90E+07	17.4790722	1.30E+04	9.4727046
	C3	C1-C2-C4	1.70E+08	18.951309	5.50E+06	15.520259
	C4	C1-C2-C3	3.50E+05	12.7656884	3.70E+04	10.518673
Category 3						
			KI	ln KI	KI	ln KI
Scenario 12	M	AF-C1	8.80E+04	11.3850921	1.40E+01	2.6390573
[P1]-P2-C1		AF-C2	2.00E+05	12.2060726	5.60E+04	10.933107
		AF-C3	6.40E+05	13.3692235	2.80E+05	12.542545
		AF-C4	2.50E+05	12.4292162	4.30E+05	12.97154
	AF	M-C1	1.30E+05	11.7752897	8.40E-12	-25.502789
		M-C2	1.20E+05	11.695247	1.80E+02	5.1929569
		M-C3	2.70E+04	10.2035921	2.00E+03	7.6009025
		M-C4	6.80E+03	8.82467789	2.70E+07	17.111347
			KI	ln KI	KI	ln KI
Scenario 13 [P1]-P2-C1-C2	M	AF-C1-C2	6.00E+06	15.60727	2.10E+01	3.0445224
		AF-C1-C3	2.00E+07	16.8112428	4.60E+01	3.8286414
		AF-C1-C4	9.80E+08	20.7030631	4.40E+03	8.3893598
		AF-C2-C3	6.00E+07	17.9098551	6.70E+07	18.020203
		AF-C2-C4	2.10E+09	21.4652032	5.30E+06	15.483217
		AF-C3-C4	1.70E+08	18.951309	1.20E+09	20.905587
	AF	M-C1-C2	1.60E+08	18.8906844	7.30E-12	-25.643147
		M-C1-C3	7.70E+07	18.159316	1.30E-13	-29.671242
		M-C1-C4	4.10E+07	17.5290826	4.40E-08	-16.939076

		M-C2-C3	4.20E+07	17.5531802	6.70E+07	18.020203
		M-C2-C4	1.50E+08	18.8261459	1.60E+07	16.588099
		M-C3-C4	1.20E+09	20.9055874	3.10E+07	17.249498
			KI	ln KI	KI	ln KI
Scenario 14 [P1]-P2-C1-C2-C3	M	AF-C1-C2-C3	4.00E+08	19.8069751	2.90E+02	5.6698809
		AF-C1-C2-C4	3.90E+10	24.3868275	1.90E+02	5.2470241
		AF-C1-C3-C4	6.40E+09	22.5795638	3.50E+03	8.1605182
		AF-C2-C3-C4	1.90E+10	23.6677048	8.60E+09	22.875028
	AF	M-C1-C2-C3	1.90E+09	21.3651197	5.50E-07	-14.413348
		M-C1-C2-C4	6.40E+09	22.5795638	5.10E-05	-9.8836849
		M-C1-C3-C4	3.60E+10	24.3067848	1.30E-08	-18.158316
		M-C2-C3-C4	7.20E+10	24.999932	6.60E+09	22.610335
			KI	ln KI	KI	ln KI
Scenario 15 [P1]-P2-C1-C2-C3-C4	M	AF-C1-C2-C3-C4	2.30E+11	26.1613451	6.10E+03	8.7160441
	AF	M-C1-C2-C3-C4	3.90E+10	24.3868275	2.70E-04	-8.2170886
			KI	ln KI	KI	ln KI
Scenario 19	M	C1-C2	4.90E+05	13.1021607	2.60E+05	12.468437
[P]-C1-C2		C1-C3	4.60E+06	15.3415669	1.70E+06	14.346139
		C1-C4	2.60E+06	14.771022	6.10E+05	13.321214
		C2-C3	3.30E+07	17.3120181	3.10E+04	10.341742
		C2-C4	1.00E+08	18.4206807	4.60E+04	10.736397
		C3-C4	2.70E+06	14.8087623	8.40E+05	13.641157
	AF	C1-C2	1.30E+07	16.3804599	8.80E-11	-23.153684
		C1-C3	1.80E+07	16.7058823	1.90E-11	-24.686582
		C1-C4	1.10E+05	11.6082356	2.60E-07	-15.162584
		C2-C3	2.30E+07	16.9510048	2.00E+04	9.9034876
		C2-C4	7.60E+06	15.8436588	9.70E+04	11.482466
		C3-C4	1.90E+07	16.7599495	1.60E+04	9.680344
			KI	ln KI	KI	ln KI
Scenario 20 [P]-C1-C2-C3	M	C1-C2-C3	4.20E+07	17.5531802	4.80E+02	6.1737861
		C1-C2-C4	5.30E+08	20.0883876	6.00E+02	6.3969297
		C1-C3-C4	2.60E+07	17.0736071	7.00E+06	15.761421

		C2-C3-C4	8.30E+08	20.5369363	1.30E+06	14.077875
	AF	C1-C2-C3	2.00E+08	19.1138279	1.30E-08	-18.158316
		C1-C2-C4	8.60E+07	18.2698579	7.30E-06	-11.827636
		C1-C3-C4	1.50E+08	18.8261459	7.80E-08	-16.366557
		C2-C3-C4	3.10E+09	21.8546679	8.40E+05	13.641157
			KI	ln KI	KI	ln KI
Scenario 21	M	C1-C2-C3-C4	6.30E+09	22.5638155	3.00E+03	8.0063676
[P]-C1-C2-C3-C4	AF	C1-C2-C3-C4	1.10E+09	20.818576	6.20E-06	-11.990961

Full Sib (00-0004)
HIS
C1-C4

Scenario	KI Range	KI Mean	ln KI Range	ln KI Mean
Scenario 1	1.50E+08	2.99E+07	18.82614585	1.52E+01
	2.40E+04		10.08580911	
Scenario 2	1.00E+09	2.09E+08	20.72326584	1.78E+01
	1.60E+05		11.98292909	
Scenario 3	9.40E+09	2.67E+09	22.96397553	2.00E+01
	7.10E+09		14.40329722	
Scenario 7	6.50E+07	1.32E+07	17.98989783	1.42E+01
	2.30E+04		10.04324949	
Scenario 8	9.40E+08	2.87E+08	20.66139043	1.75E+01
	3.50E+05		12.76568843	
Scenario 12	6.40E+05	1.83E+05	13.36922346	1.15E+01
	6.80E+03		8.824677891	
Scenario 13	2.10E+09	4.17E+08	21.46520318	1.86E+01
	6.00E+06		15.60727003	
Scenario 14	7.20E+10	2.26E+10	24.99993196	2.30E+01
	4.00E+08		19.80697511	
Scenario 15	2.30E+11	1.35E+11	26.16134515	2.53E+01
	3.90E+10		24.38682748	
Scenario 19	1.00E+08	1.87E+07	18.42068074	1.57E+01
	1.10E+05		11.60823564	
Scenario 20	3.10E+09	6.21E+08	21.85466795	1.92E+01
	2.60E+07		17.0736071	
Scenario 21	6.30E+09	3.70E+09	22.56381547	2.17E+01
	1.10E+09		20.81857602	

Half Sib (00-0033)
HIS

C1*-C4

Scenario	KI Range	KI Mean	ln KI Range	ln KI Mean
Scenario 1	1.40E+08	6.41E+06	18.75715298	5.58E+00
	9.60E-10		-20.76408783	
Scenario 2	1.00E+08	6.65E+06	18.42068074	3.47E+00
	4.60E-09		-19.19720953	
Scenario 3	6.70E+06	9.34E+05	15.71761808	5.53E+00
	5.40E-08		-16.73428179	
Scenario 7	2.20E+06	4.76E+05	14.60396792	7.49E+00
	3.50E-02		-3.352407217	
Scenario 8	5.50E+06	1.39E+06	15.52025865	8.92E+00
	1.20E+00		0.182321557	
Scenario 12	2.70E+07	3.47E+06	17.11134742	5.44E+00
	8.40E-12		-25.50278941	
Scenario 13	1.20E+09	1.16E+08	20.90558739	4.11E+00
	1.30E-13		-29.67124194	
Scenario 14	8.60E+09	1.90E+09	22.87502804	2.76E+00
	1.30E-08		-18.15831648	
Scenario 15	6.10E+03	3.05E+03	8.71604405	2.49E-01
	2.70E-04		-8.217088599	
Scenario 19	1.70E+06	3.02E+05	14.34613881	3.58E+00
	1.90E-11		-24.68658214	
Scenario 20	7.00E+06	1.14E+06	15.76142071	1.21E+00
	1.30E-08		-18.15831648	
Scenario 21	3.00E+03	1.50E+03	8.006367568	-1.99E+00
	6.20E-06		-11.99096127	

Half Sib (00-0033)			FALSE	TRUE	TRUE
Scenario 1	Missing/Test	Ref Combo	ln KI	ln KI*	KI
[C]-P-C1	C1*	M-C2	4.317488114	11.41861479	9.10E+04
		M-C3	9.61580548	11.41861479	9.10E+04
		M-C4	8.5754621	11.41861479	9.10E+04
		AF-C2	-19.62465355	5.480638923	2.40E+02
		AF-C3	-18.55994281	6.086774727	4.40E+02
		AF-C4	-20.76408783	5.886104031	3.60E+02
	C2	M-C1*	0	7.170119543	1.30E+03
		M-C3	9.472704636	9.472704636	1.30E+04

	M-C4	11.27720313	11.27720313	7.90E+04
	AF-C1*	-1.609437912	6.253828812	5.20E+02
	AF-C3	13.12236338	13.12236338	5.00E+05
	AF-C4	7.901007052	7.901007052	2.70E+03
C3	M-C1*	7.090076836	6.253828812	5.20E+02
	M-C2	11.23848862	11.23848862	7.60E+04
	M-C4	14.771022	14.771022	2.60E+06
	AF-C1*	2.014903021	9.392661929	1.20E+04
	AF-C2	15.70258021	15.70258021	6.60E+06
	AF-C4	6.65929392	6.65929392	7.80E+02
C4	M-C1*	4.86753445	7.696212639	2.20E+03
	M-C2	11.8493977	11.8493977	1.40E+05
	M-C3	13.55414579	13.55414579	7.70E+05
	AF-C1*	8.748304912	17.72753356	5.00E+07
	AF-C2	18.75715298	18.75715298	1.40E+08
	AF-C3	14.94691267	14.94691267	3.10E+06

APPENDIX C

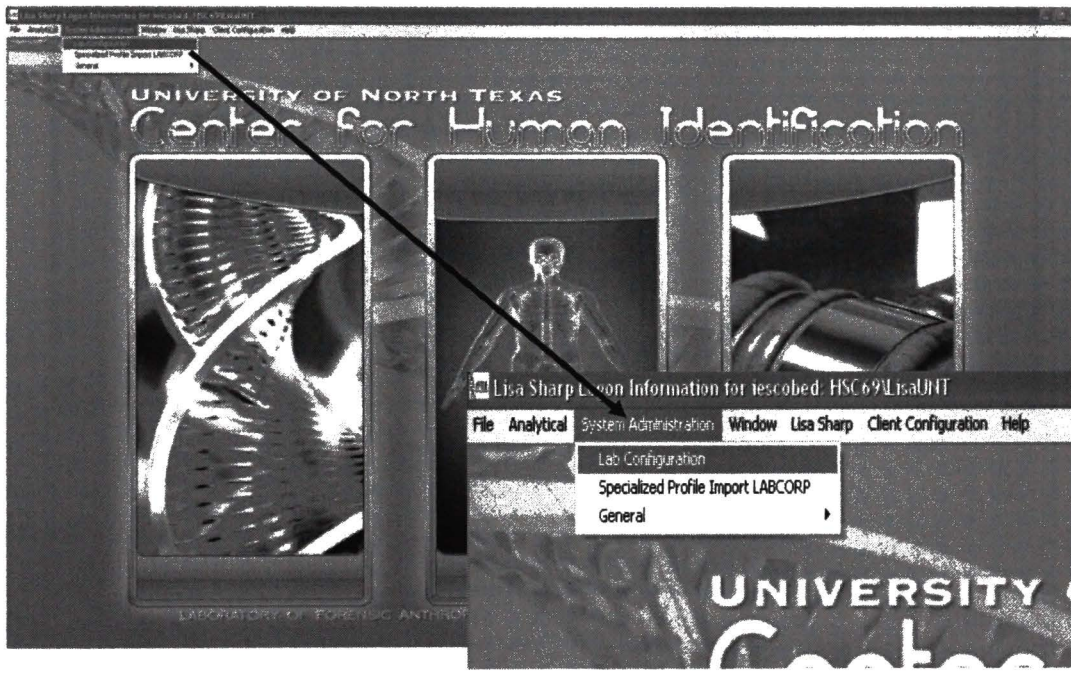
PROTOCOL FOR USING LISA FOR KINSHIP ANALYSIS

Instructions for using LISA for Kinship Analysis

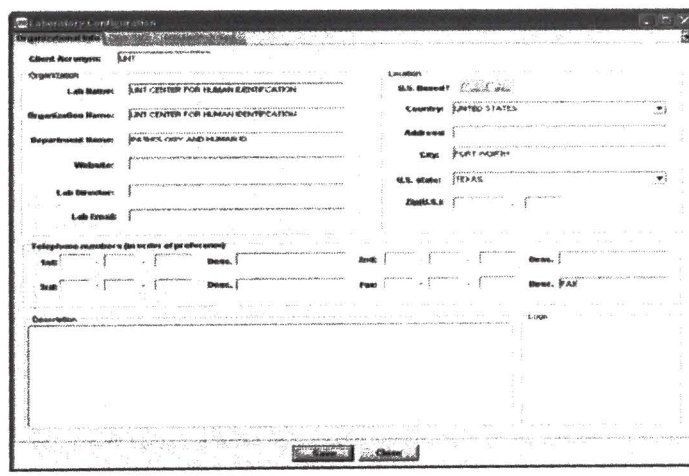
I. Making a New Case Set:

1. Locate **FTI: LISA Analytical** icon on desktop and click it.

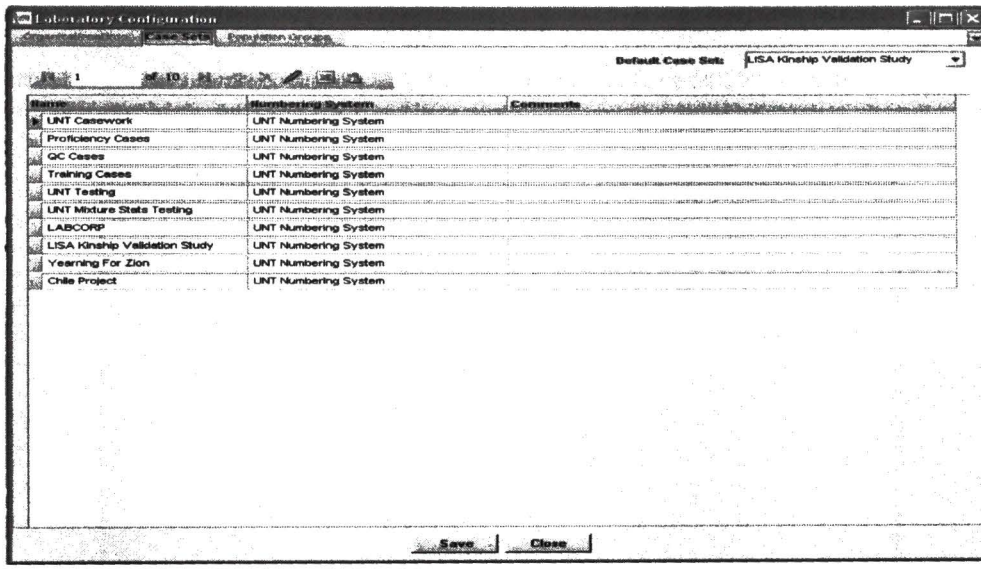
2. Once the program is open, locate and click the **System Administration** tab. On top of the menu select the **Lab Configuration** tab.

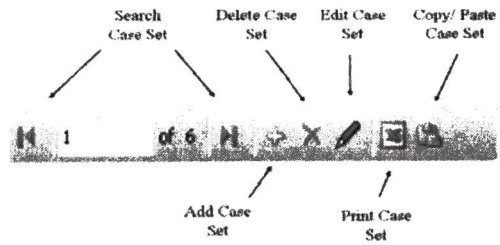


3. The **Lab Configuration** window will appear. On top of the window select the **Case Sets** tab to add a case set into LISA.

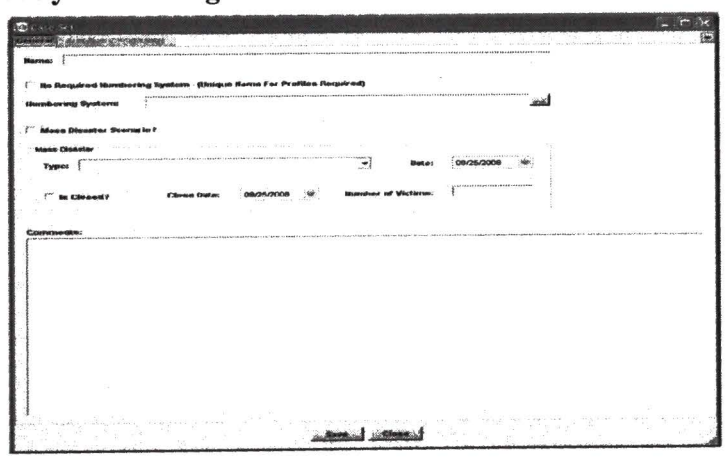


4. The **Case Sets** tab contains several icons to choose from in order to enter, delete, edit, and search for case sets. It also contains a list of currently saved case sets saved in LISA. A default case set can be selected for use in LISA by using the **Default Case Set** scroll menu.



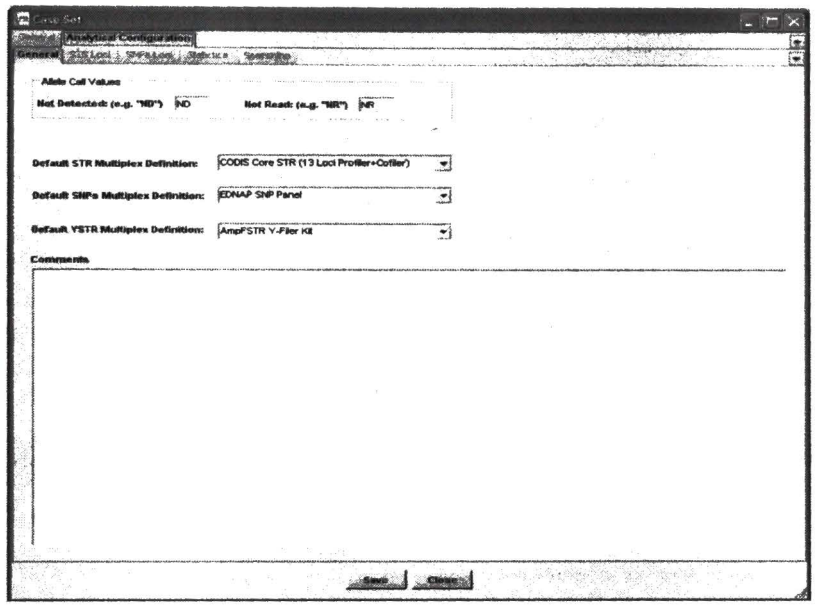


5. Click the yellow plus icon add a case set. The **Case Set** window will appear with two tabs **General** and **Analytical Configuration.**



6. In the **General** tab, type in the name of the case set, the number of victims, and any comments regarding with the case set in the required fields. Select the numbering system, and start/closing dates for the case set. Check the required boxes if the case set does not require a numbering system, if the case set is a mass disaster scenario, or if the case is closed.

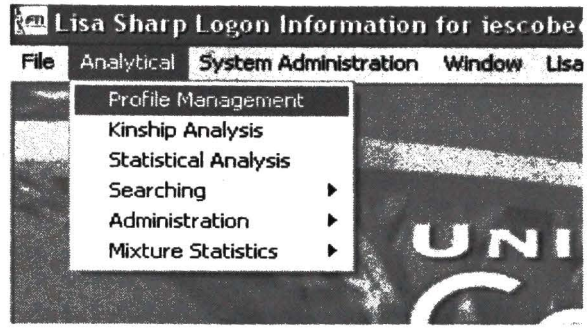
7. In the **Analytical Configuration** tab, set up the population databases and statistical/searching settings for the **Case Set.**



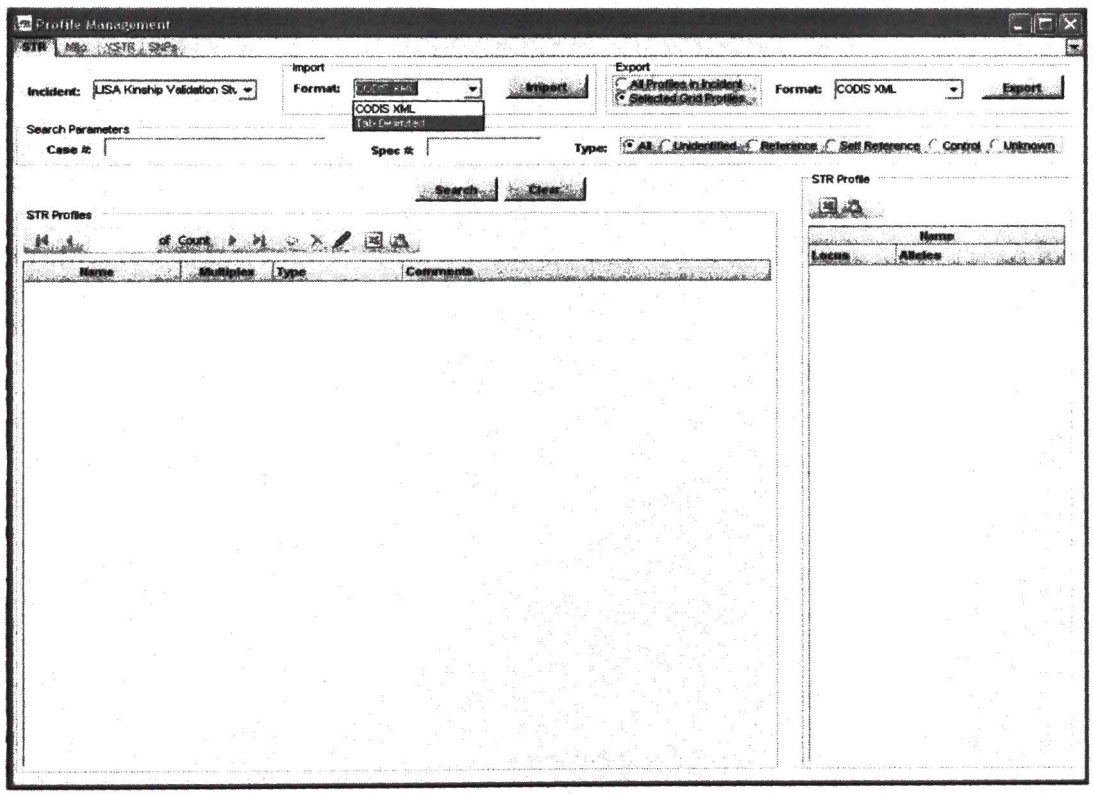
8. Click the **Save** tab, once all of the information has been filled out to save the case set. Once the case set has been saved, the file will be seen in the list of case sets found in the **Case Sets** section of the **Laboratory Configuration** window.

II. Importing DNA Profiles into a Case Set

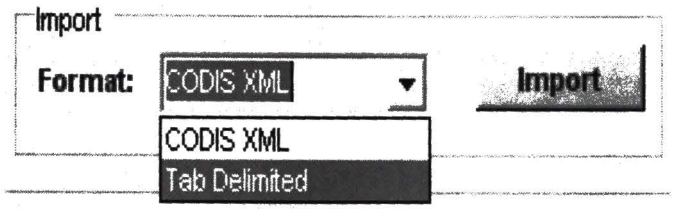
1. To import DNA profiles to a particular case set, go to the main window of the program and click the **Analytical** tab. Select **Profile Management.**



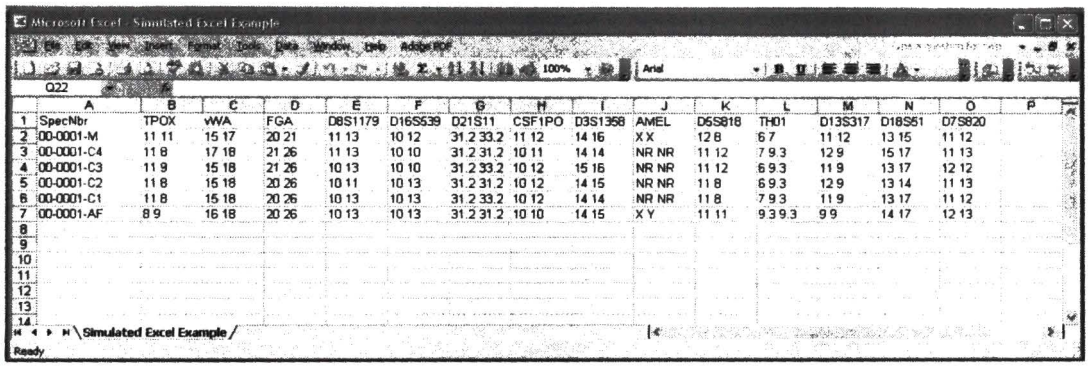
2. In the top left corner of the **Profile Management** window select one of the tabs for the type of data you will be importing (**STR, Mito, Y-STR, or SNP**). Next select the case set where profiles will be imported to using the **Incident** scroll menu at the left corner of the window.



3. To import DNA profiles saved in a particular format into LISA, use the **Format** scroll menu in the **Import** section to select the **CODIS XML** or **Tab Delimited** formats. To import profiles from CODIS, Genotyper, or GeneMapper ID select the **CODIS XML** format option. After the selection is made, click the **Import** tab.

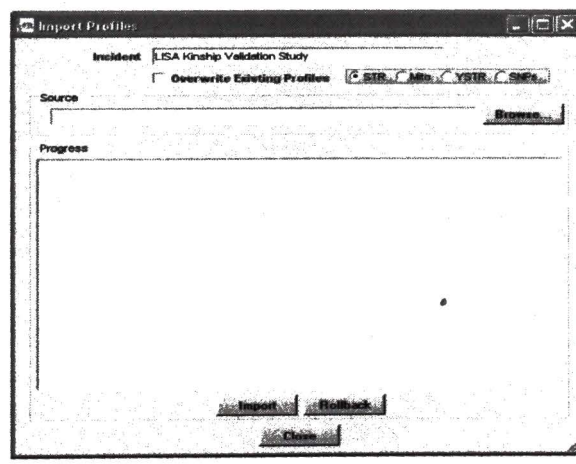


An excel example of profiles in Tab Delimited format

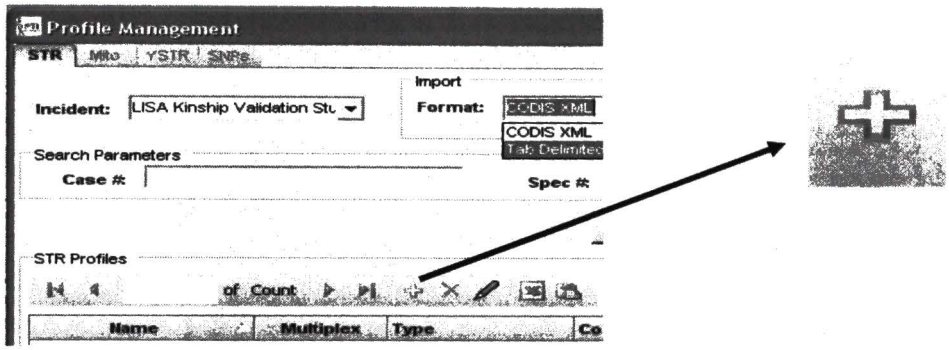


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	SpecNbr	TPOX	vWA	FGA	D8S1179	D16S539	D21S11	CSF1PO	D3S1358	AMEL	D5S818	TH01	D13S317	D18S51	D7S820	
2	00-0001-M	11 11	15 17	20 21	11 13	10 12	31.2 33.2	11 12	14 16	X X	12 8	6 7	11 12	13 15	11 12	
3	00-0001-C4	11 8	17 18	21 26	11 13	10 10	31.2 31.2	10 11	14 14	NR NR	11 12	7 9.3	12 9	15 17	11 13	
4	00-0001-C3	11 9	15 18	21 26	10 13	10 10	31.2 33.2	10 12	15 16	NR NR	11 12	6 9.3	11 9	13 17	12 12	
5	00-0001-C2	11 8	15 18	20 26	10 11	10 13	31.2 31.2	10 12	14 15	NR NR	11 8	6 9.3	12 9	13 14	11 13	
6	00-0001-C1	11 8	15 18	20 26	10 13	10 13	31.2 33.2	10 12	14 14	NR NR	11 8	7 9.3	11 9	13 17	11 12	
7	00-0001-AF	8 9	16 18	20 26	10 13	10 13	31.2 31.2	10 10	14 15	X Y	11 11	9.3 9.3	9 9	14 17	12 13	
8																
9																
10																
11																
12																
13																
14																

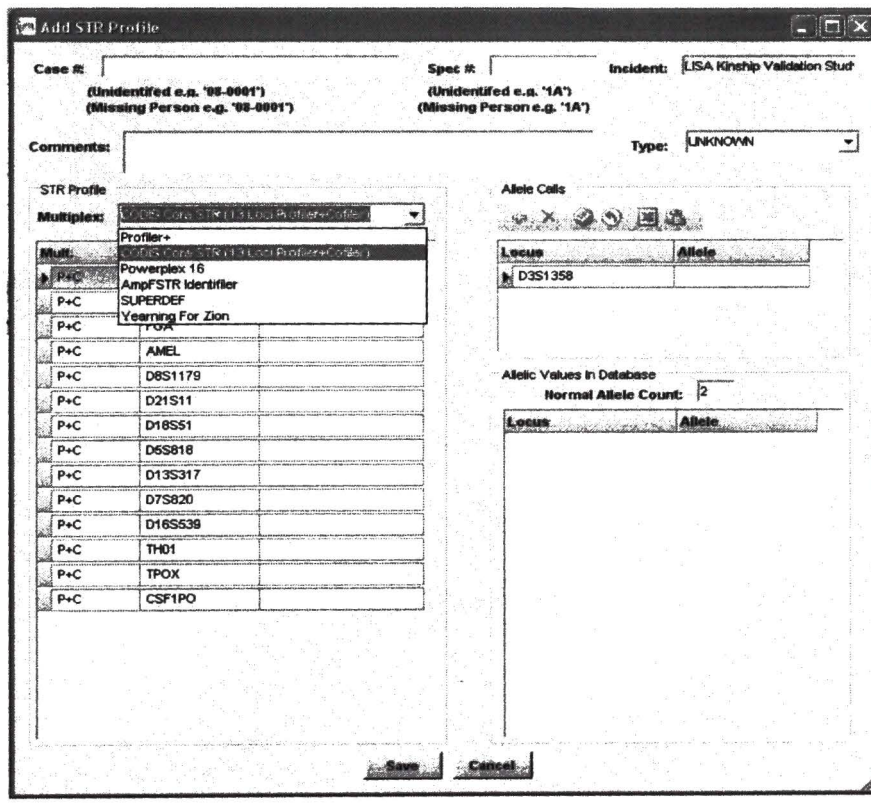
4. An **Import Profiles** window will appear on the screen where you can browse for the format file that will be imported into LISA. Select the type of DNA profiles that will be imported into LISA like STRs, Mito, Y-STRs, or SNPs. Check the **Overwrite Existing Profiles** box to overwrite existing profiles in LISA. Click the **Browse** tab to search for the format file that will be imported. Once selected, click the **Import** tab or to cancel import click the **Rollback** tab.



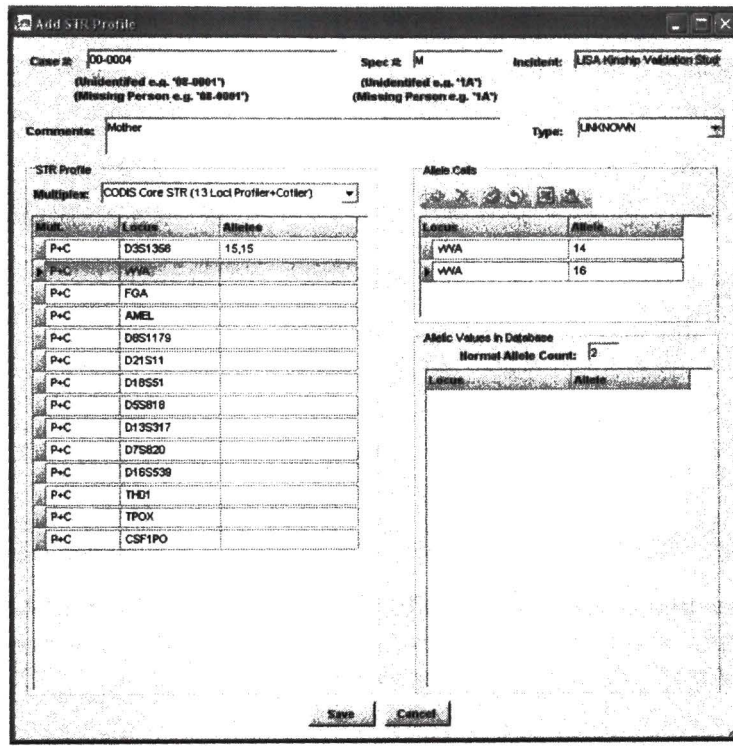
5. Importing DNA profiles can also be done manually. Click the yellow plus icon found in the **STR** section in order to add a blank STR DNA profile to the case set.



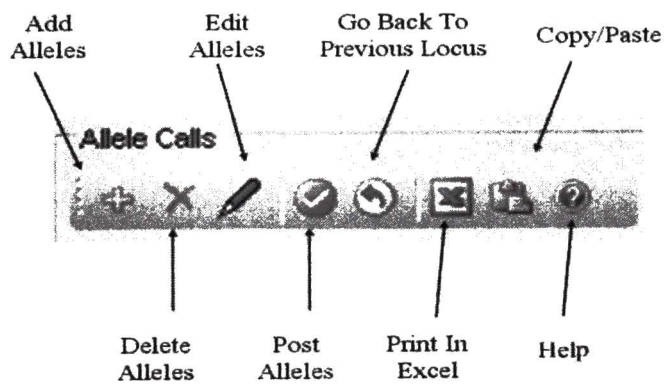
6. The **Add STR Profile** window will open and it will contain several fields that need to be filled out like the case number, specimen number, and any comments that goes with the profile. Remember the case number must be in a 00-0000 format. In the **Type** scroll menu select the type of profile (reference, unknown, mixture, control, etc). At the **STR Profile** section look for the **Multiplex** scroll menu to select the mutiplex system used on the sample.



7. In the **STR Profile** section, click any of the loci found on the list. Add the allele calls for the selected locus in the **Allele Call** section on the right side of the window. When the first allele is typed in its field, press the **Enter** button, add the next allele call, and press **Enter** again to add the next allele call for the next locus. For a homozygote locus, you must type identical numbers in both fields. Loci with no alleles type in NR (Not Read) or ND (Not Detected) in both fields.

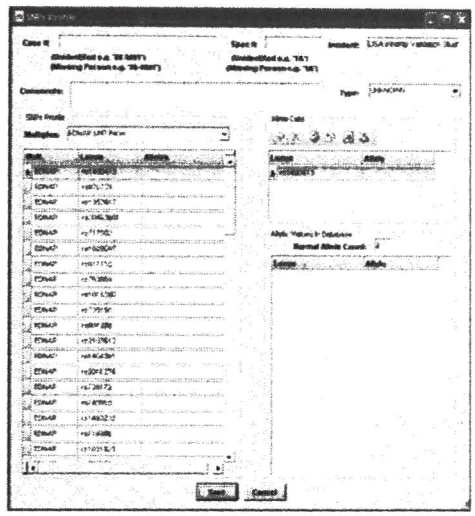


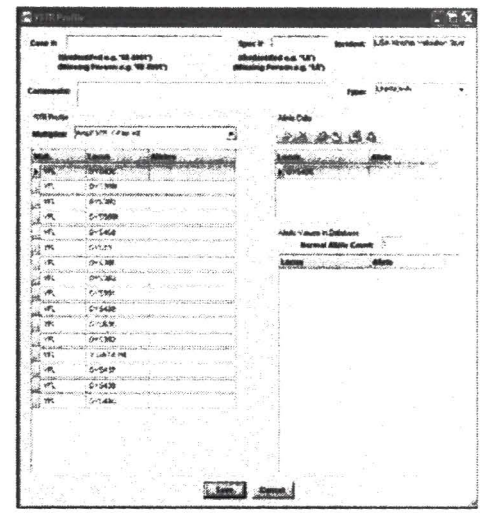
8. In the **Allele Calls** section there are several icons to select for the addition, deletion, and editing of allele calls as well as other capabilities



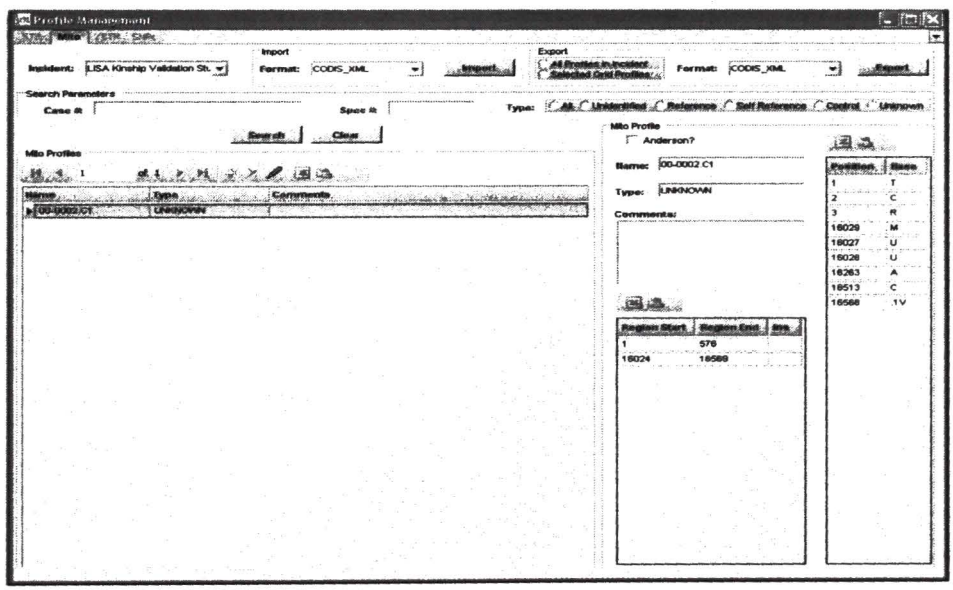
9. Once the entire profile has been typed in, click the **Save** tab at the bottom of the window or to cancel click the **Cancel** tab.

10. The entire process is the same for entering Y-STR and SNP profiles when selecting the **Y-STR** and **SNP** tabs of the **Profile Management** window.

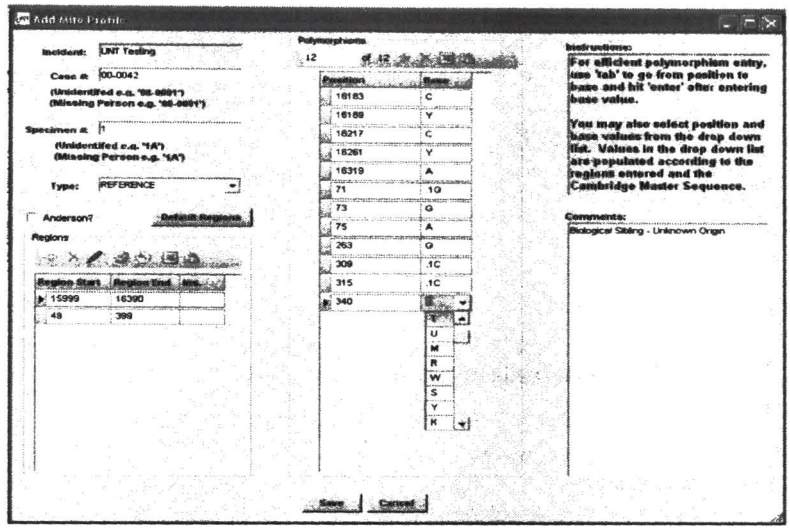




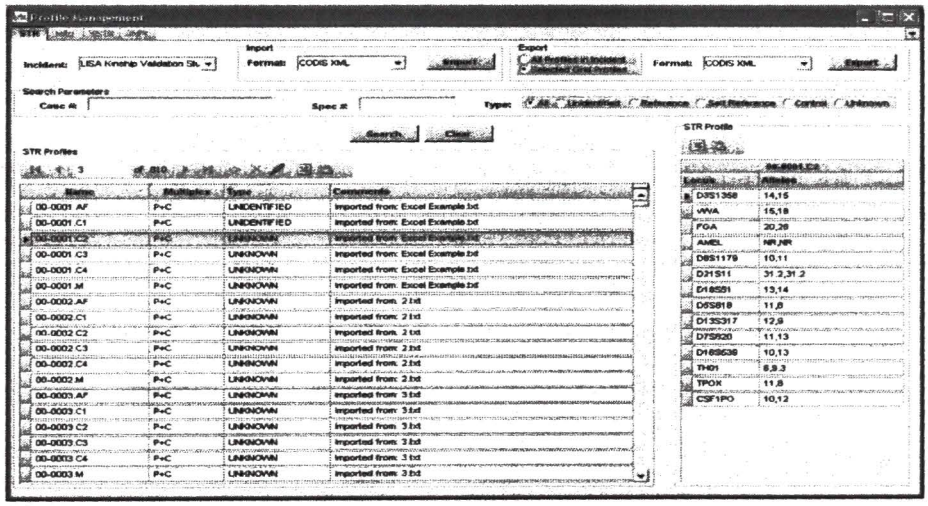
11. For entering Mito profiles, click the **Mito** tab of the **Profile Management** window. Look for the yellow plus icon to add Mito profile. Mito profiles can also be imported in XML formats from CODIS and Sequencher in the **Import** section of the **Profile Management** window.



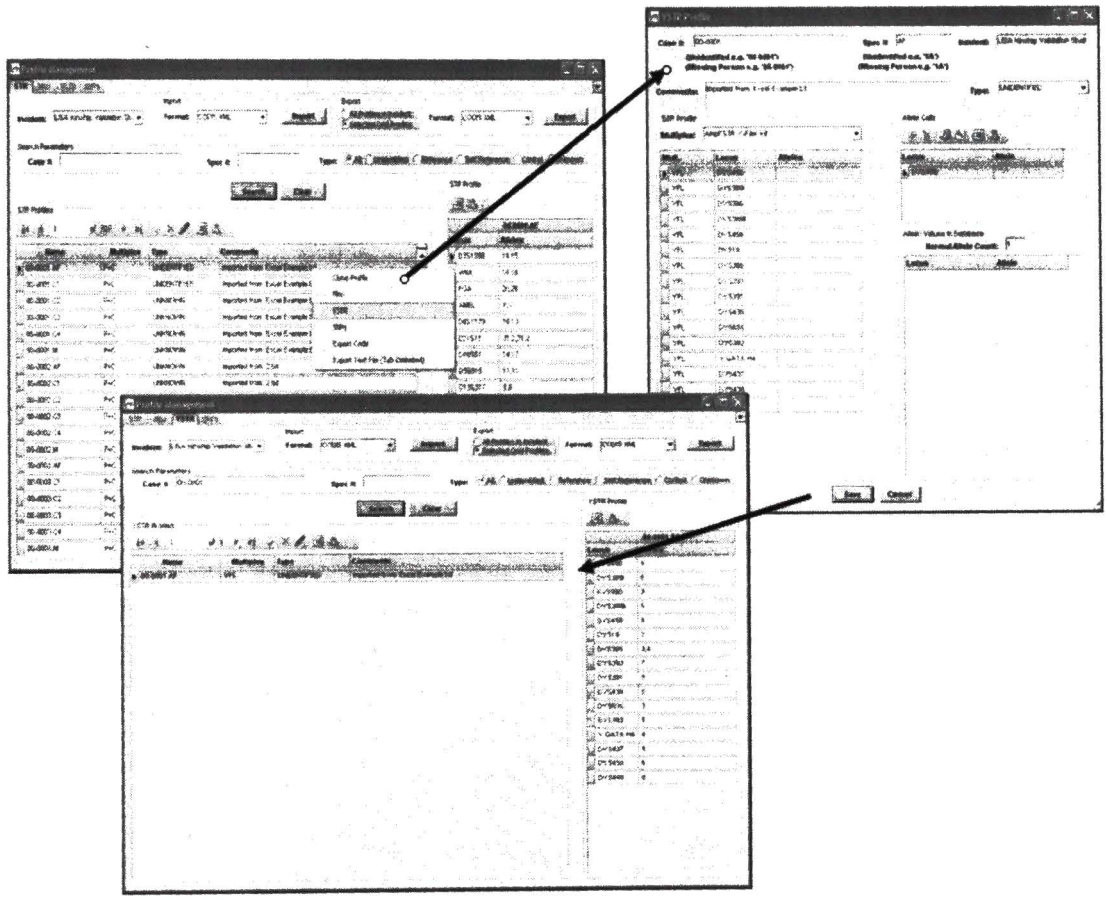
12. The **Add Mito Profile** has fields to put in the case/specimen numbers, and comments. It also contains a scroll menu for selecting the sample type. Click the **Default Regions** to enter the data for the profile. Type in or select the position and its corresponding base or letter symbol for heteroplasmy using the scroll menus in **Position** and **Base** columns. For insertions type .1N and for deletions type in -. Some instructions for entering Mito profiles can be found in the **Instructions** section on the right top corner of the window. Once all the information is entered in, click the **Save** tab.



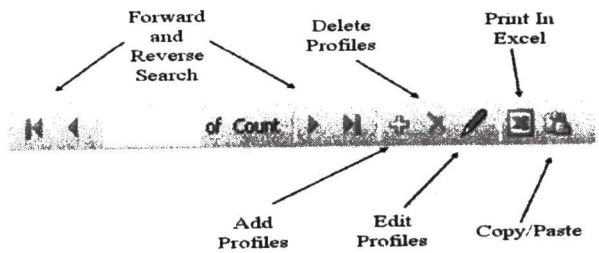
13. All STR, Mito, Y-STR, and SNP profiles added to the case set in LISA can be found in the corresponding **STR, Mito, Y-STR,** and **SNP** windows. To preview a particular profile select the profile from the list and the profile contents can be seen on the right side of the **Profile Management** window in the corresponding **STR, Mito, Y-STR,** and **SNP Profile** section.



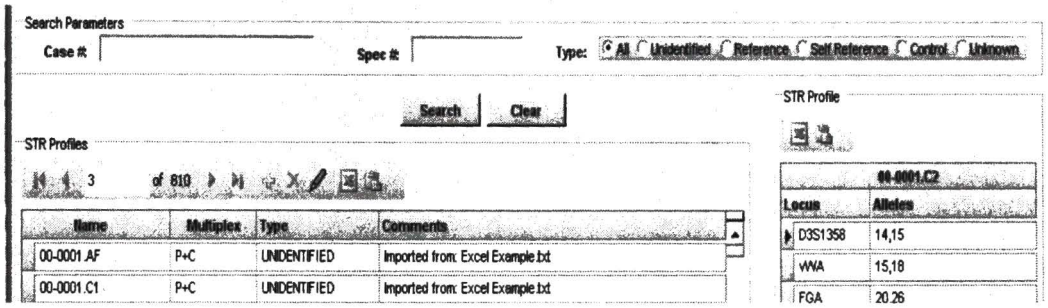
14. To manually import additional Mito, Y-STR, and SNP profiles for a saved STR profile in LISA just right click at a highlighted profile of interest. Select the type of genetic data that will be added in and an **Add Mito, Y-STR,** or **SNP Profile** window will appear to input the information. Click **Save** and the result is an STR profile with a Mito, Y-STR, or SNP profile with the same name. The same process can be done vice versa when adding STR profiles.



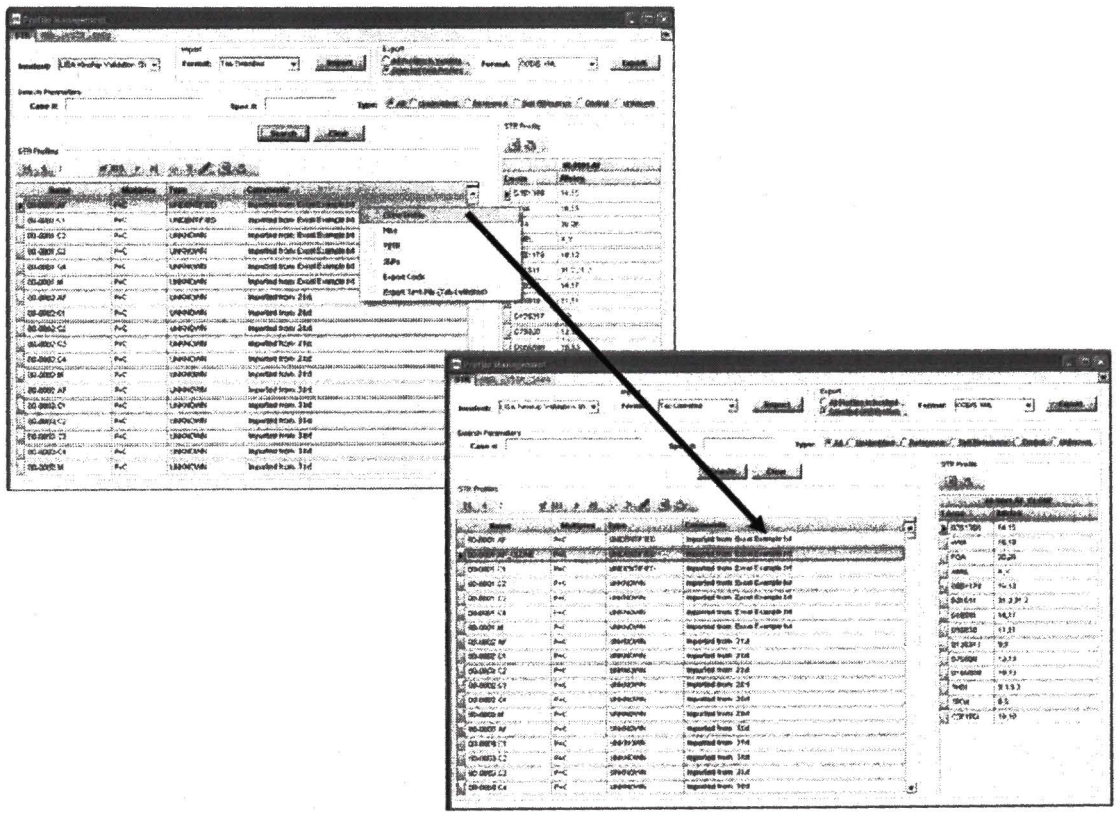
15. Each **STR, Mito, Y-STR,** and **SNP** windows will have a set of icons for the search, addition, deletion, editing, printing, and copying/pasting of DNA profiles in the case set.



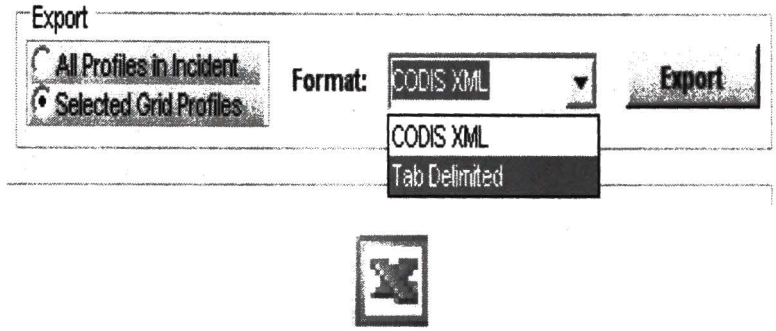
16. To search for DNA profiles in any of the **STR, Mito, Y-STR,** and **SNP** windows is by typing the case or specimen number of the profile in the require field, or click the **Search** tab in the **Search Parameters** section. The search can be narrow down by sample type by clicking sample type in the **Type** section. The resulting profile list can be organized by the sample name, multiplex, type, or comments by clicking their corresponding tabs.



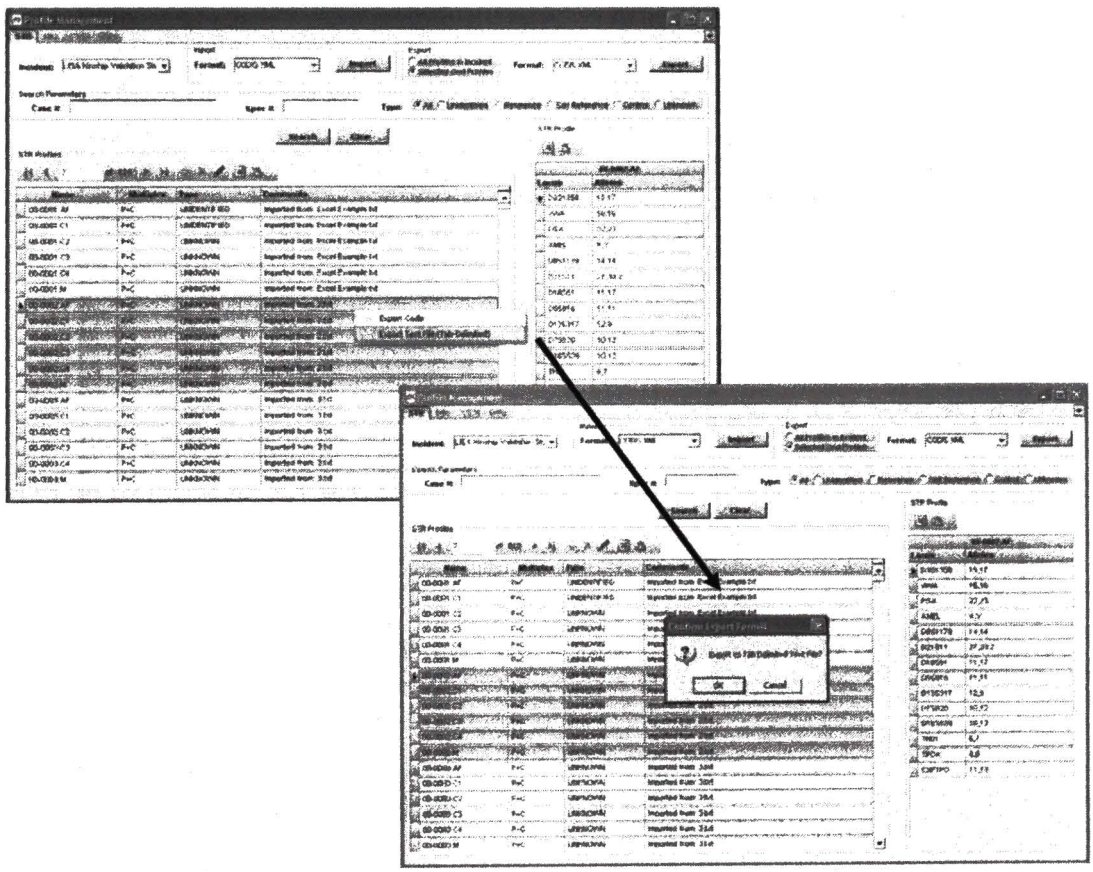
17. To clone or duplicate a DNA profile for editing, right click the profile in the list and select the **Clone Profile** option. Either an **Add STR, Mito, Y-STR,** and **SNPs Profile** window will appear where it can be edited. Click **Save**, and the cloned profile will appear on the profile list of the **Profile Management** window.



18. To export all or select profiles from the case set in a particular format, go to the **Export** section. Select **All Profiles in Incident** or **Selected Grid Profiles** and select any program that will be used from the **Format** scroll menu. To export profiles to an excel format just click the excel icon.

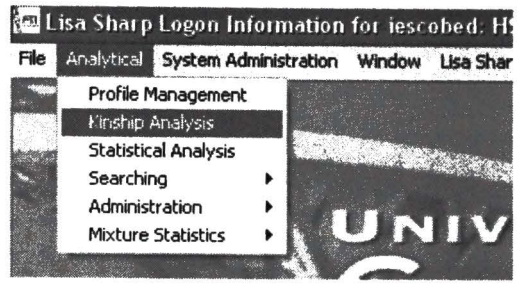


19. Another way to export profiles is by highlighting the selected profiles and right click on the list. Select the file format that the profiles will be exported in. Click the **Ok** or the **Cancel** tab in the resulting **Confirm Export Format** window.

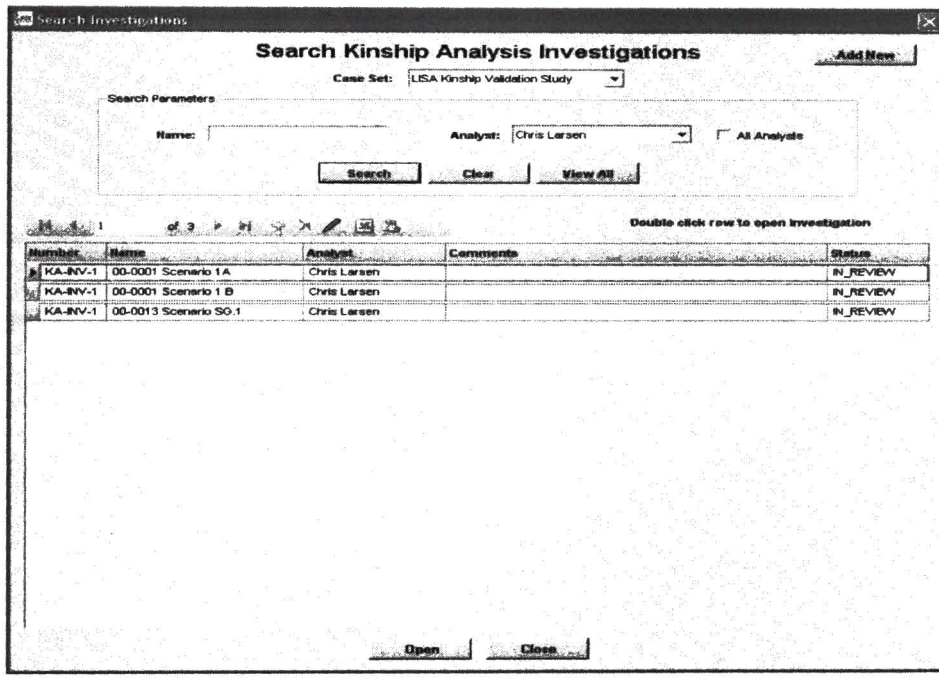


III. Creating and Searching Investigation Files for Kinship Analysis

1. To create an investigation file for kinship analysis, go to the main window, click the **Analytical** tab, and select **Kinship Analysis.**

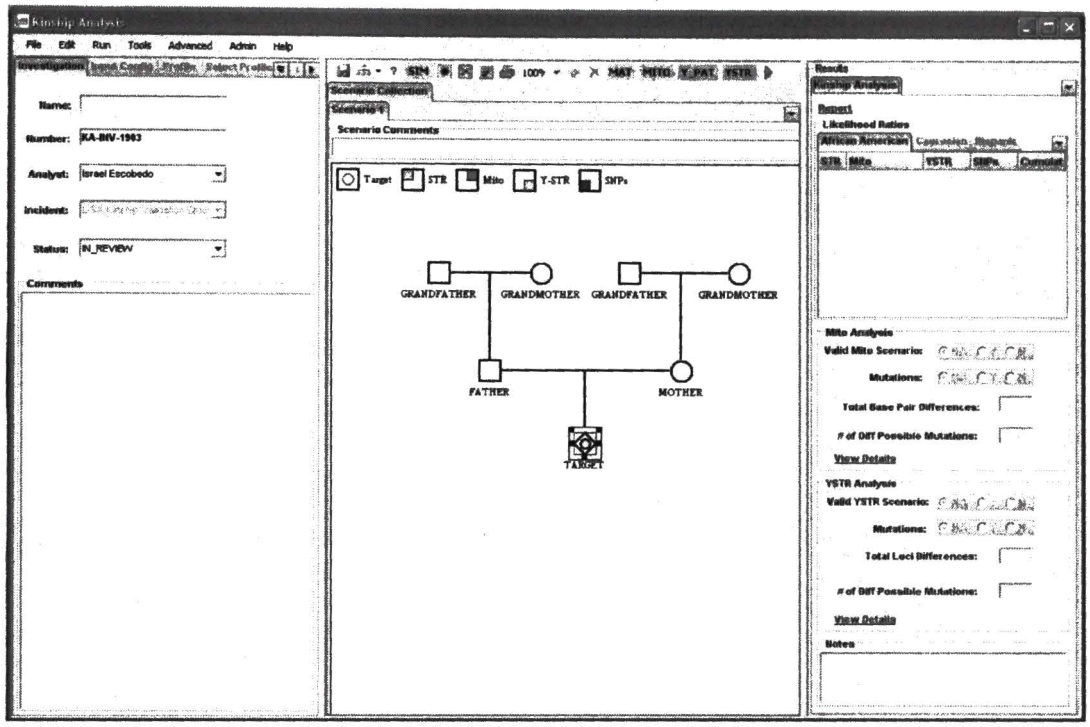


2. A **Search Investigation** will open containing icons for searching, adding, deleting, and editing investigation files for kinship analysis. Use the **Case Set** scroll menu to select the case set where the investigations files will be added to.

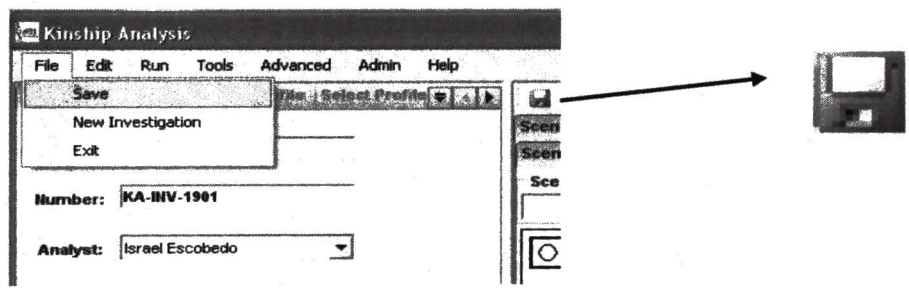


3. To create a new investigation file for kinship analysis click the **Add New** tab at the top right corner of the **Search Investigations** window or the yellow plus icon.

4. A **Kinship Analysis** window will open, locate the **Investigation** tab section. Type in the name of the investigation file, and assign an investigation file number. Select the analyst's name and the status of the investigation in the required fields.

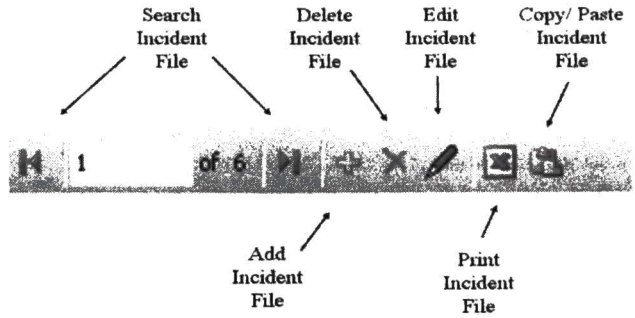


5. Once all of the required information is filled out, go to **File,** and select the **Save** tab to save the investigation file or click the **Save** icon.



6. To check or search any of the saved investigation files, first close the **Kinship Analysis** window. To search for an investigation file, select the case set using the **Case Set** scroll menu or check the **Use All Incidents** box. Next you type in the name of the investigation file in the **Name** field at the **Search Parameters** section of the **Search Investigations** window. Select the analyst in the **Analyst** scroll menu to narrow down search. Next click the **Search** tab in to complete the search. Click **Clear** tab to clear search results.

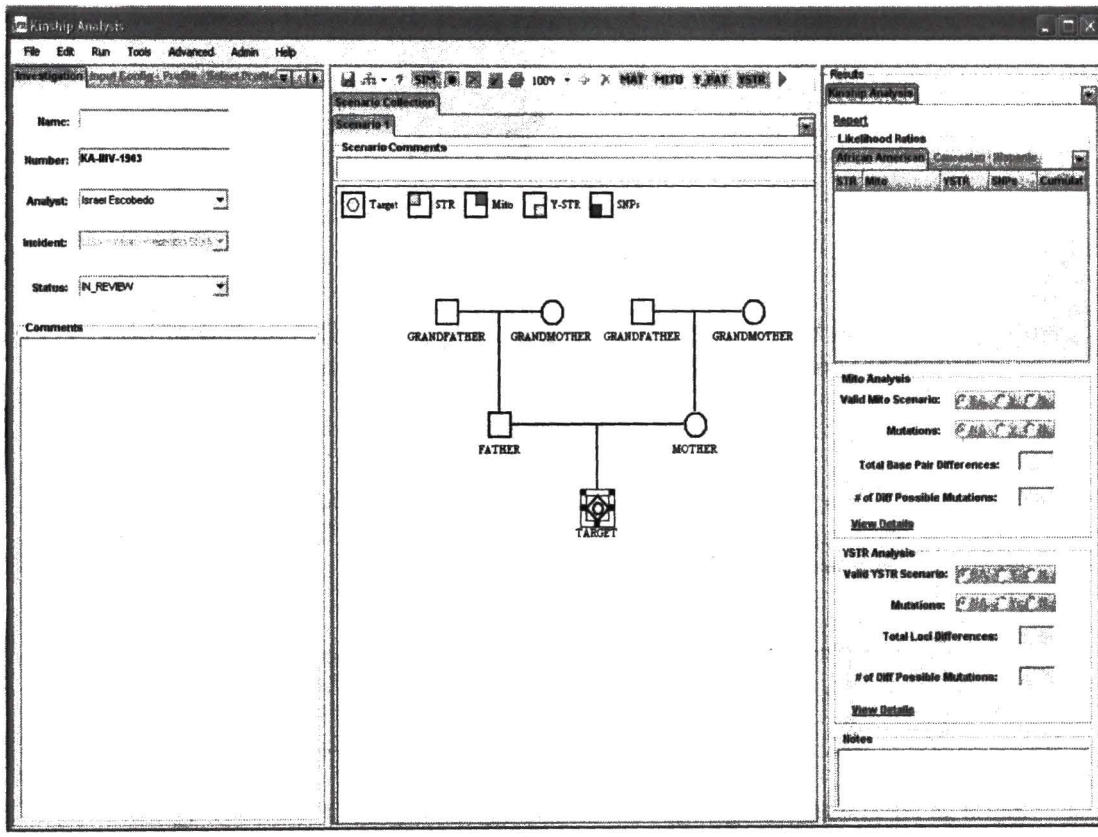
7. To delete, add, edit, print, copy/paste, or search through a list of investigation files click to the appropriate icons located below the **Search Parameters** section.



IV. Constructing Pedigrees

1. Double click the investigation file that will be used for constructing pedigrees in the **Search Investigation** window.

2. A **Kinship Analysis** window will open containing three major sections. The middle section with the pedigree is the area where pedigree construction is performed.

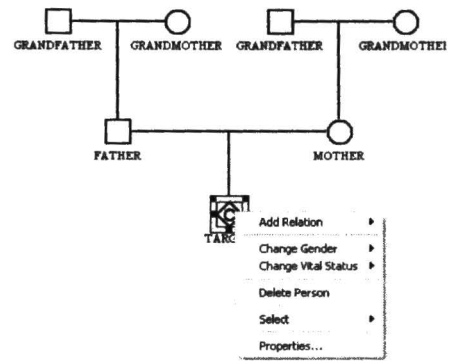


3. Type in the name of the pedigree or scenario in the **Scenario Comments** field at the **Scenario Collection** section. Each investigation file has a limit of 14 scenarios or pedigree.

4. Three ways exist to construct pedigrees. Modification, Manual, and Duplicating.

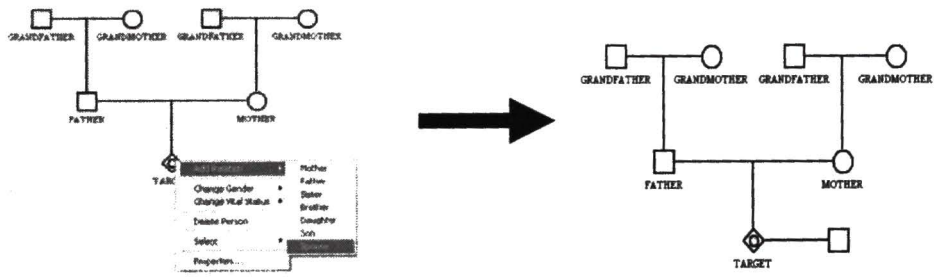
5. Modification

To modify the template pedigree provided by LISA, left click any individual in the pedigree to highlight it and right click to obtain scroll menu. The scroll menu contains many options to add/delete individuals, select individuals, or change sex/vital status of the highlighted individual.



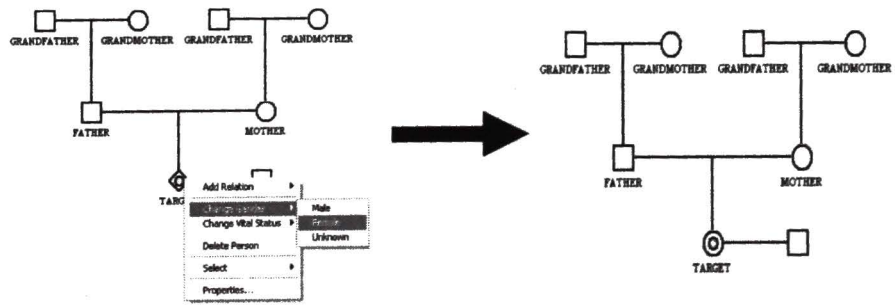
A) **Add Relation**: Adds any type of relative to the highlighted individual. Options include adding a mother, father, sister, brother, daughter, son, or spouse.

For example: Add spouse to the highlighted target.



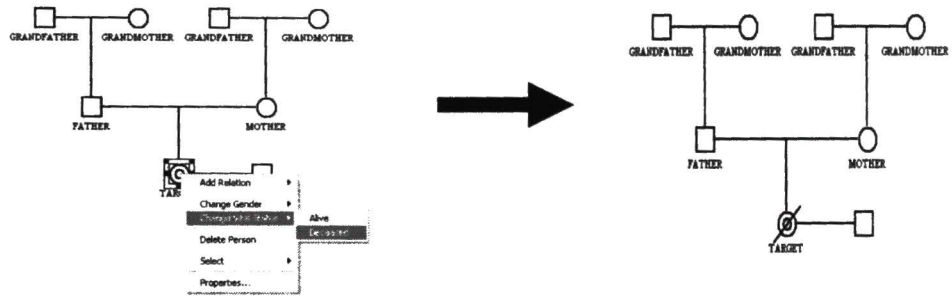
B) **Change Gender**: Changes the gender of the highlighted individual to male, female, or unknown.

For example: Change the gender of the highlighted target from unknown to female.



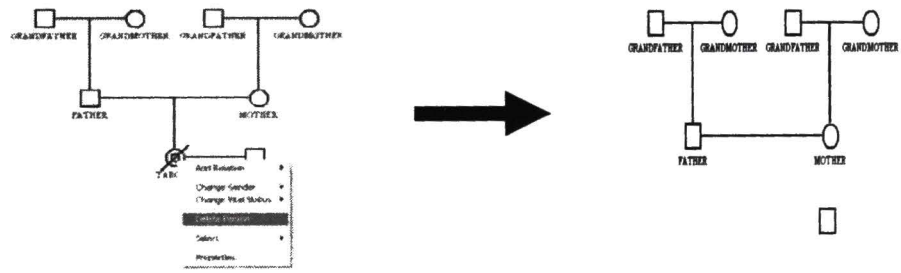
C) **Change Vital Status**: Change the vital status of the highlighted individual.

For example: Change the alive status of the highlighted target to deceased.



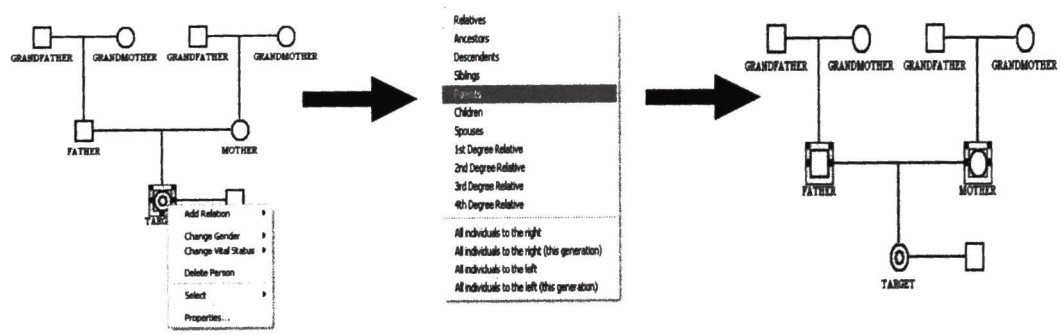
D) **Delete Person**: Deletes highlighted target.

For example: Delete highlighted target.



E) **Select**: Highlights certain individual types within the pedigree.

For example: Select parents.

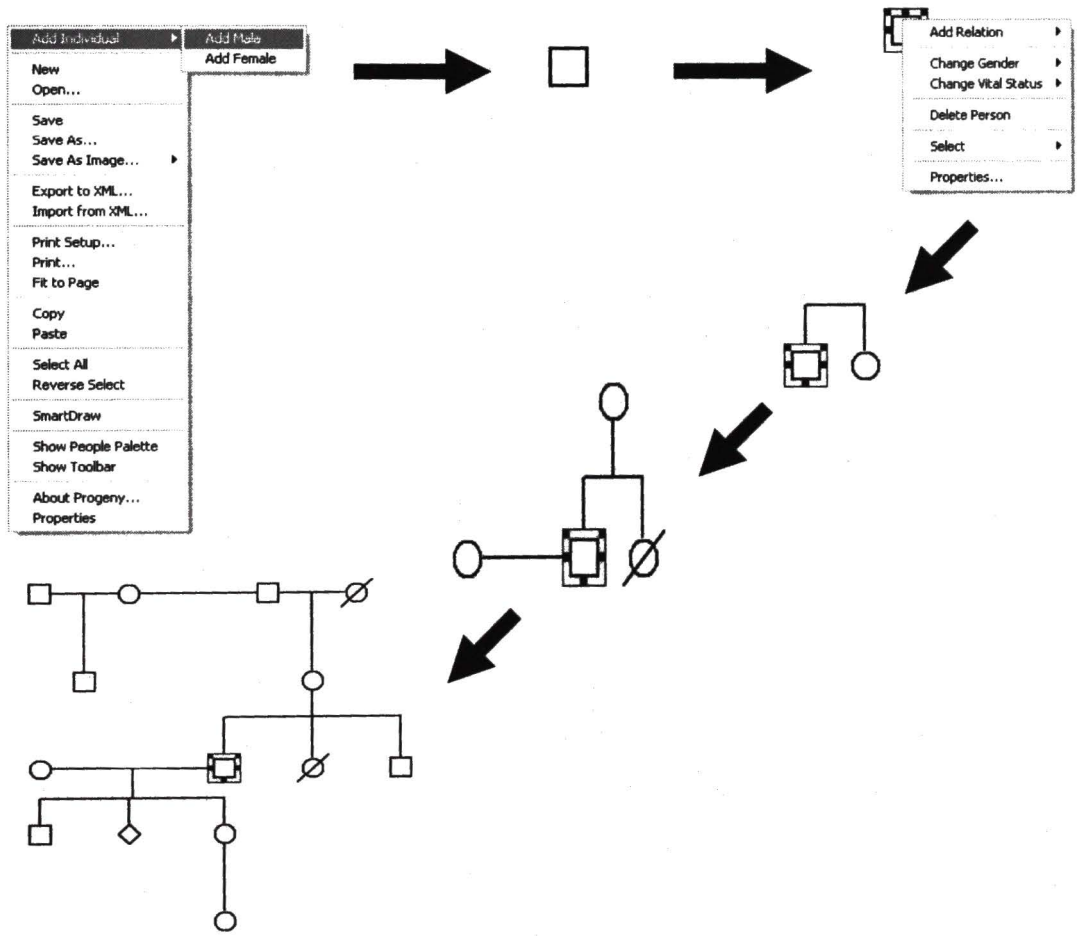


If there is a need to start over the pedigree just click the **Template Pedigrees** icon at the top of the **Kinship Analysis** window to replace current pedigree with a new template pedigree.

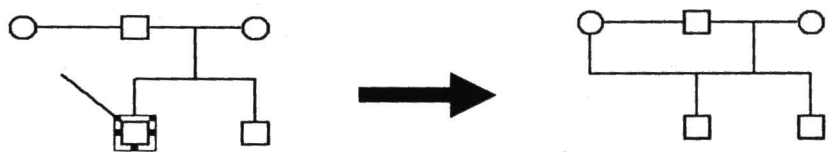
6. **Manual**

Construct a pedigree manually by deleting the template pedigree using the **Clear Pedigree** icon located at the top of the **Kinship Analysis** window or right click/drag to highlight the pedigree and press the **Delete** button.

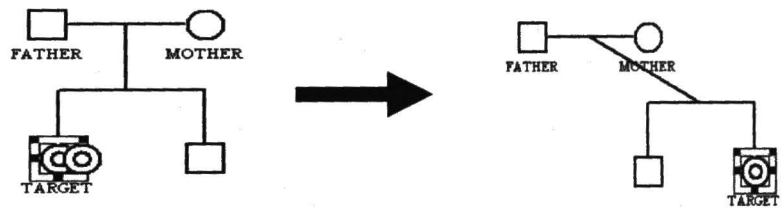
Next right click in the white space to obtain a scroll menu. Select the **Add Individual** to add male or female individual symbol to build the pedigree. Highlight the individual and right click to add relatives, and change gender or vita status. From there on the pedigree can gradually be constructed into a more complex pedigree.



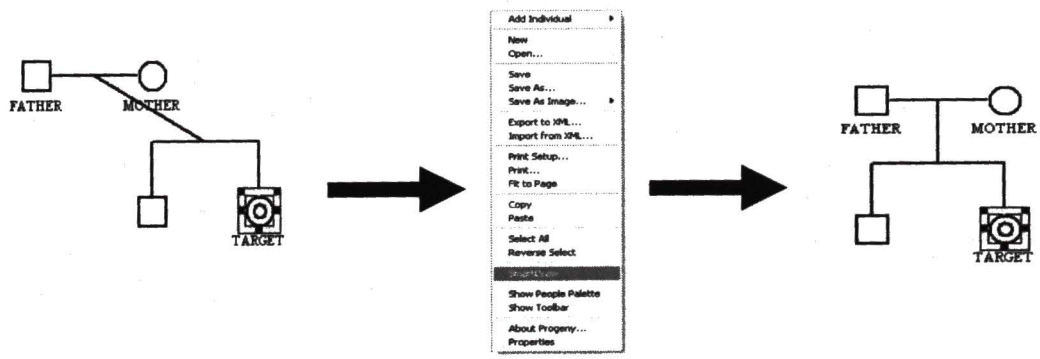
To make new associations between individuals within the pedigree just highlight the desired individual, click the black squares on the yellow frame, and drag the resulting black line to center of another individual. Always click on the black square that is positioned to the direction of the individual that will be part of the new relationship. In doing this you can create incest relationships or other types of familial relationships.



To rearrange the pedigree into a different pedigree is by highlighting the individual and dragging that individual to the direction where the change will be made.



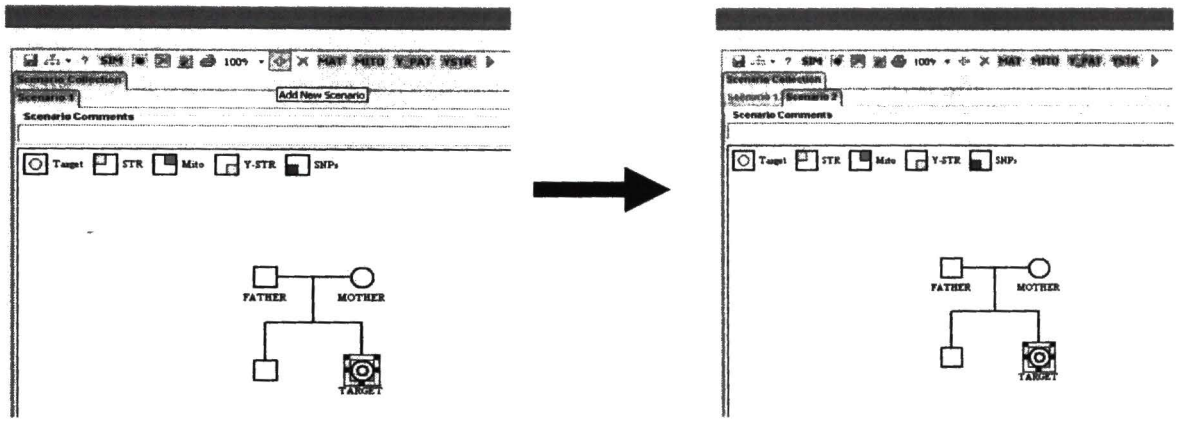
To straighten out the pedigree into a correct alignment or symmetrical pedigree just right click in the white space to get the scroll menu to select **Smart Draw**.



To move the pedigree around, just highlight the pedigree and drag drop anywhere in the white area of the widow.

9. **Duplicating**

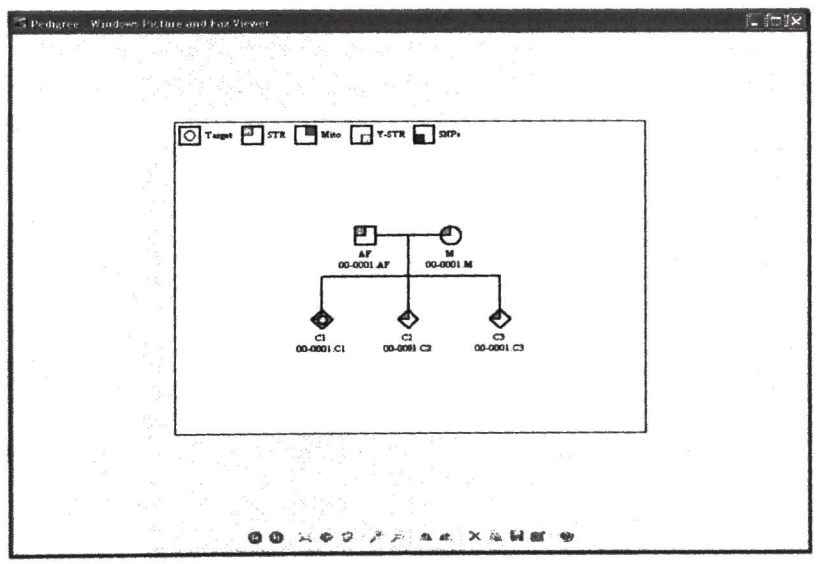
To duplicate a pedigree, just add scenarios to the current pedigree using the yellow plus sign icon and it will duplicate the pedigree exactly in all the scenarios that were added.



Another way to duplicate a pedigree is highlighted, copy, and paste in a blank scenario.

10. To save pedigrees just click the **Save** icon.

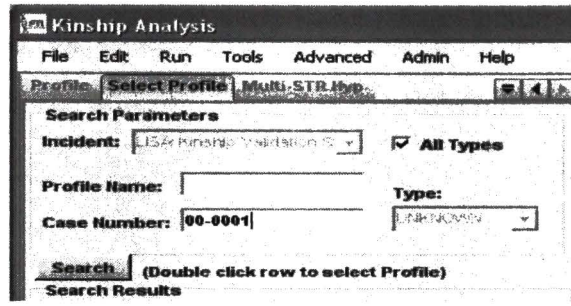
11. To export pedigree in a JPEG, BMP, GIF, PNG, or TIFF. Right click, select **Save As Image**, and select desire format.



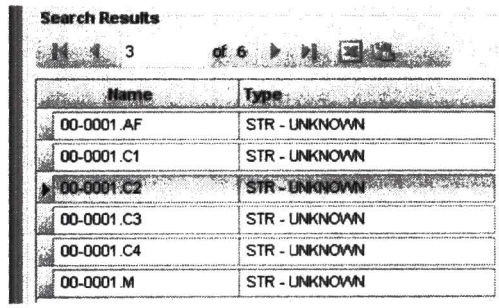
V. Importing DNA Profiles onto a Pedigree

1. At the left side of the **Kinship Analysis** window, select the **Search Profile** tab to select the DNA profiles that will be imported on to the pedigree.

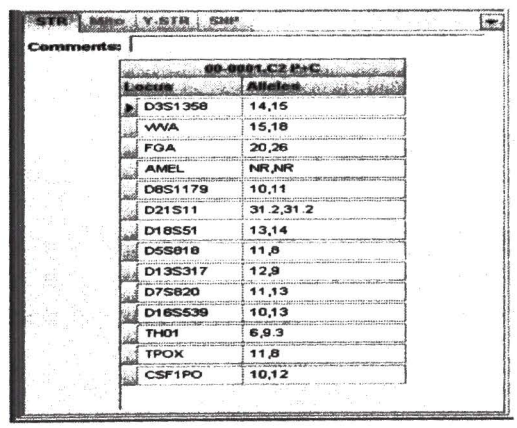
2. In the **Select Parameters** section, enter the profile name or case number in the required fields. To find a type of profile uncheck the **All Types** box to select the profile type in the **Type Scroll** menu. Once the information has been entered, click the **Search** tab.



3. The results of the search are listed in the **Search Results** section below the **Select Parameters** section.



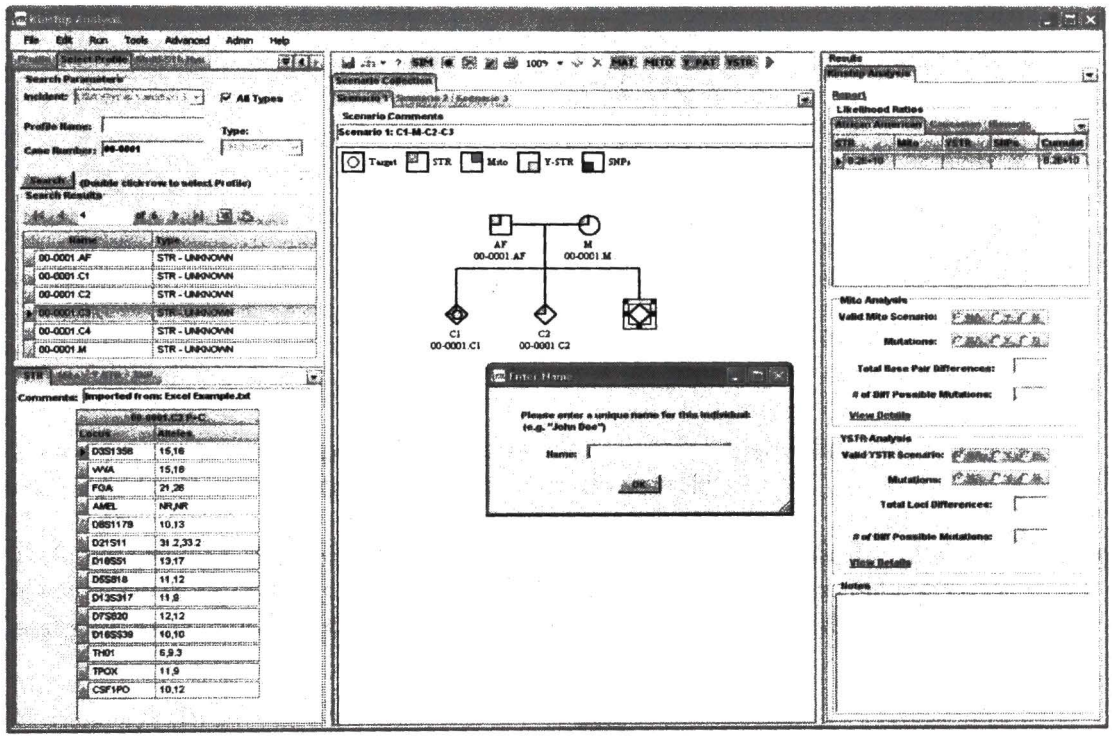
4. To view the DNA profile of a sample, double click any sample on the list in the **Search Results** section. The profile of the selected sample will be shown below the Search Results section.



5. To view the Mito, Y-STR, or SNP profiles for a sample just click the **Mito, Y-STR, or SNP** tabs above contents of the profile. Same thing can be done when looking back STR profiles using the **STR** tab.

6. To import DNA profiles in to a pedigree, select or highlight any individual in the pedigree, and double click the sample DNA profile that will be added to the pedigree.

7. An **Enter Name** window will appear, here you enter the name or label the individual pedigree. To rename individuals in the pedigree select **Tools** at the upper left corner and select **Assign Name**. The **Enter Name** window appears to change the name.

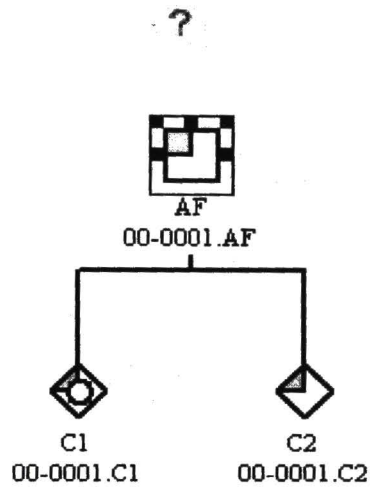


8. Same thing can be done with Mito, Y-STR, and SNP profile. For every type of profile added to an individual in a pedigree it will be color coded.

9. To delete DNA profiles from individuals of the pedigree, highlight the individual with the DNA profile and select the icon with black X in a green box at the top of the **Kinship Analysis** window to delete the desired profile.

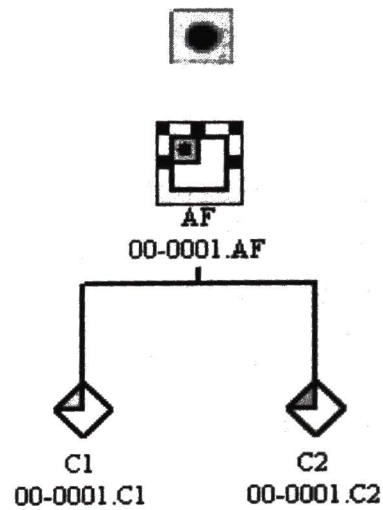
VI. Performing Kinship Analysis

1. Once all DNA profiles are imported to the pedigree, select or highlight the target individual in the pedigree that will be tested for kinship analysis. At the top of the **Kinship Analysis** window, click the icon with the question mark in the yellow circle to assign the hypothesis to the selected target. A yellow circle will be seen in the selected individual in the pedigree.

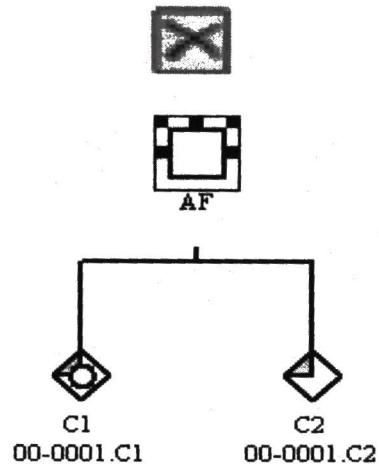


2. To remove the assigned hypothesis to selected individual, highlight targeted individual and click the icon with a question mark in a yellow circle.

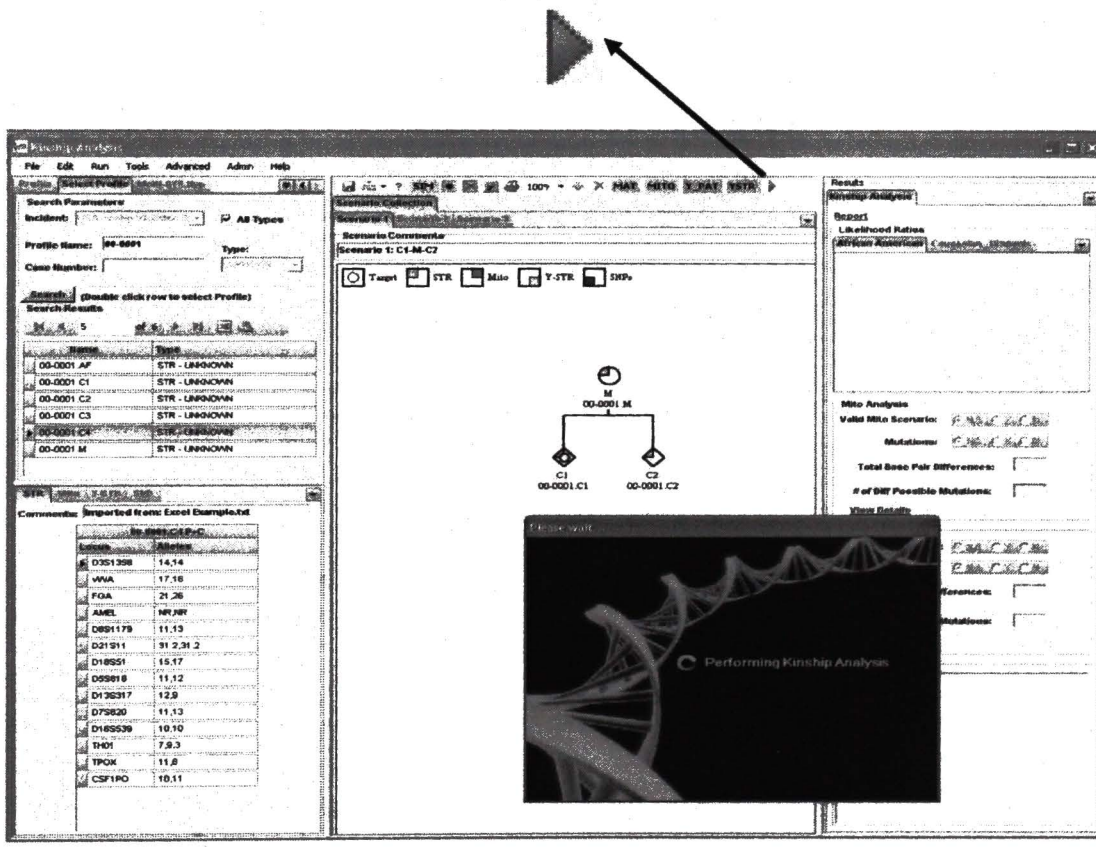
3. To block DNA profiles of certain individuals in the pedigree for kinship analysis used the black circle in a green box icon. A black dot will appear on the selected individual in the pedigree. To remove block just highlight individual and click the block icon.



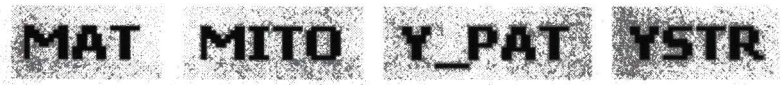
4. Another way to exclude DNA profiles from kinship analysis is to select the individual and click delete profile icon.



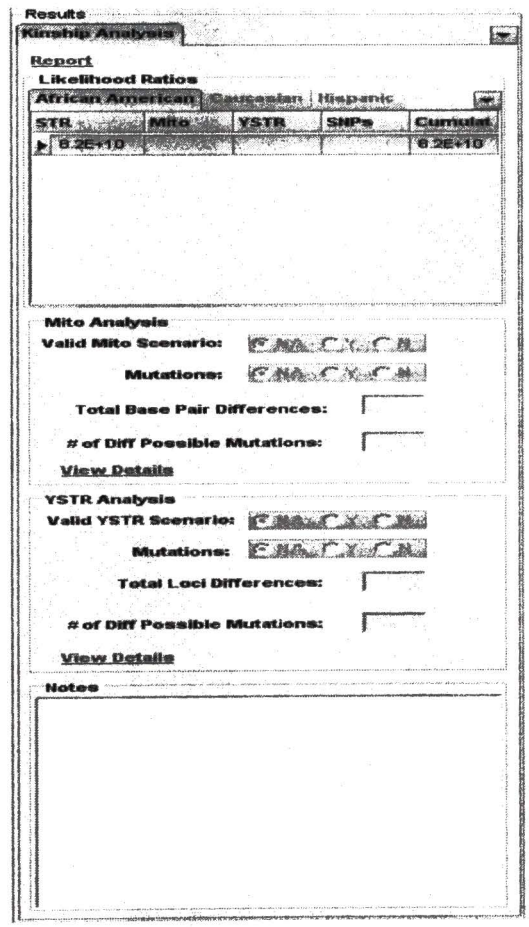
5. To perform kinship analysis, click the green arrow icon at the top of the **Kinship Analysis** window. A **Please Wait** window will appear during the kinship analysis.



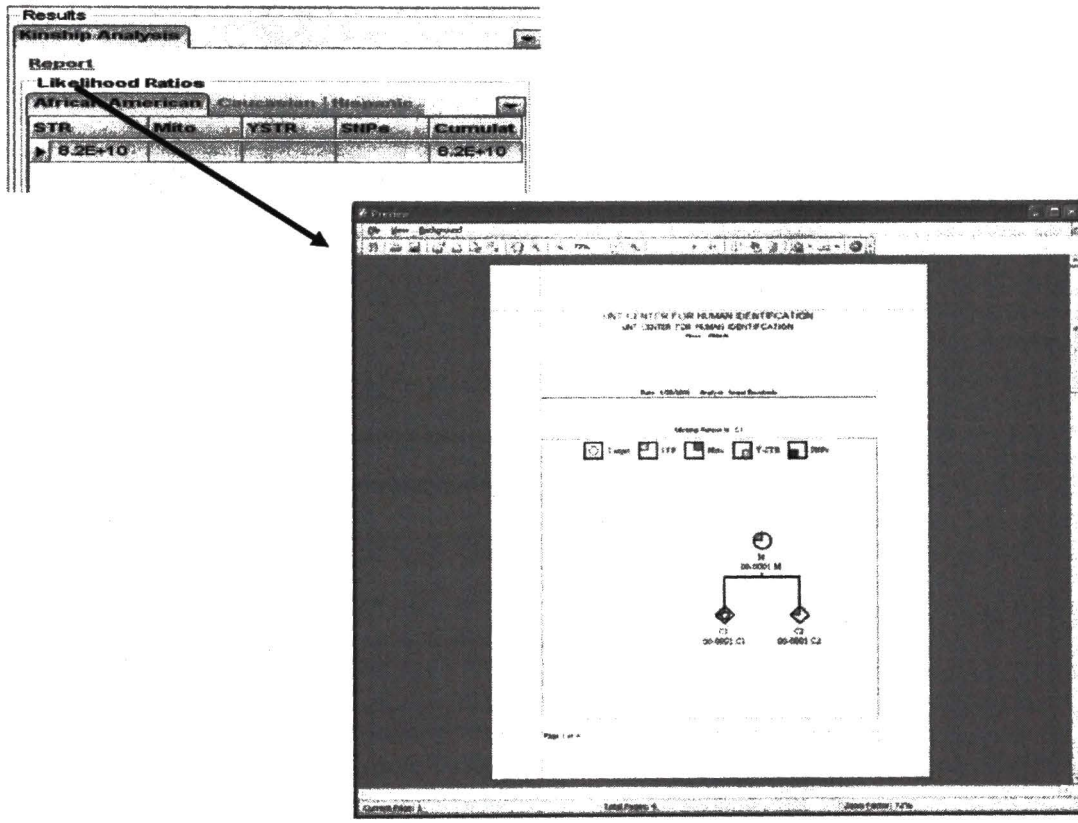
6. Kinship analysis for only Mito or Y-STR can be done by clicking the pink **Mito** tab or the blue **Y-STR** tab at the top **Kinship Analysis** window. The **MAT** and **Y_PAT** tabs mark the either maternal or paternal relatives of the pedigrees based on Mito and Y-STR profiles.



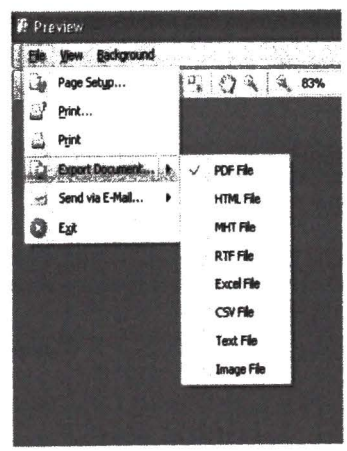
7. The likelihood ratio results of the kinship analysis are shown at the right side of the Kinship Analysis window. Likelihood ratios for STR, Mito, Y-STR, SNP, and cumulative are shown for African Americans, Caucasians, and Hispanics. Other result information Mito and Y-STR results can be also seen below the **Likelihood Ratio** section. A **Notes** section exists below **Y-STR Analysis** section to write notes about results. Kinship analysis uses Charles Brenner's DNA·VIEW Program.



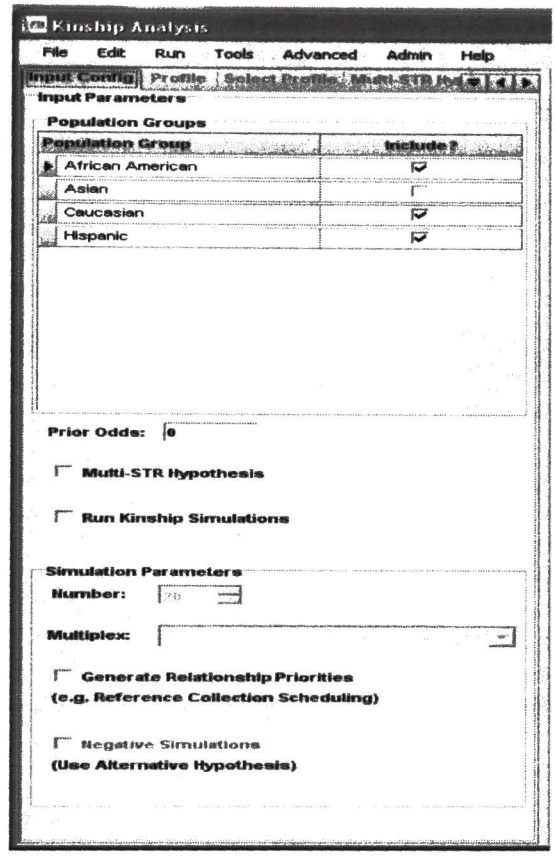
8. Click **Report** to obtain a printable four page report for the results of the kinship analysis. The Report will include a picture of the pedigree, the profiles of the tested/missing individual and individuals from the pedigree, the allelic frequencies used, the equations used, date, and the name of the analyst.



9. The report can be printed, and saved in any format like a pdf file by using the **File** scroll menu in the left side of the **Preview** window.

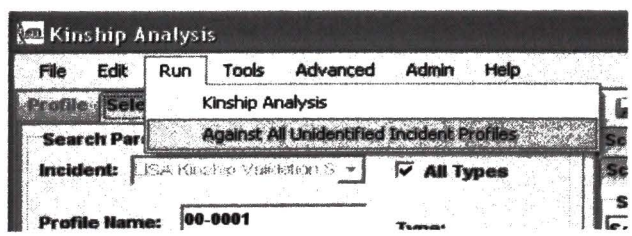


10. Modifications for kinship analysis can be done on the left side of the **Kinship Analysis** window under **Input Config** tab. In the section, populations can be selected, prior odds entered, select simulation parameters, and much more.

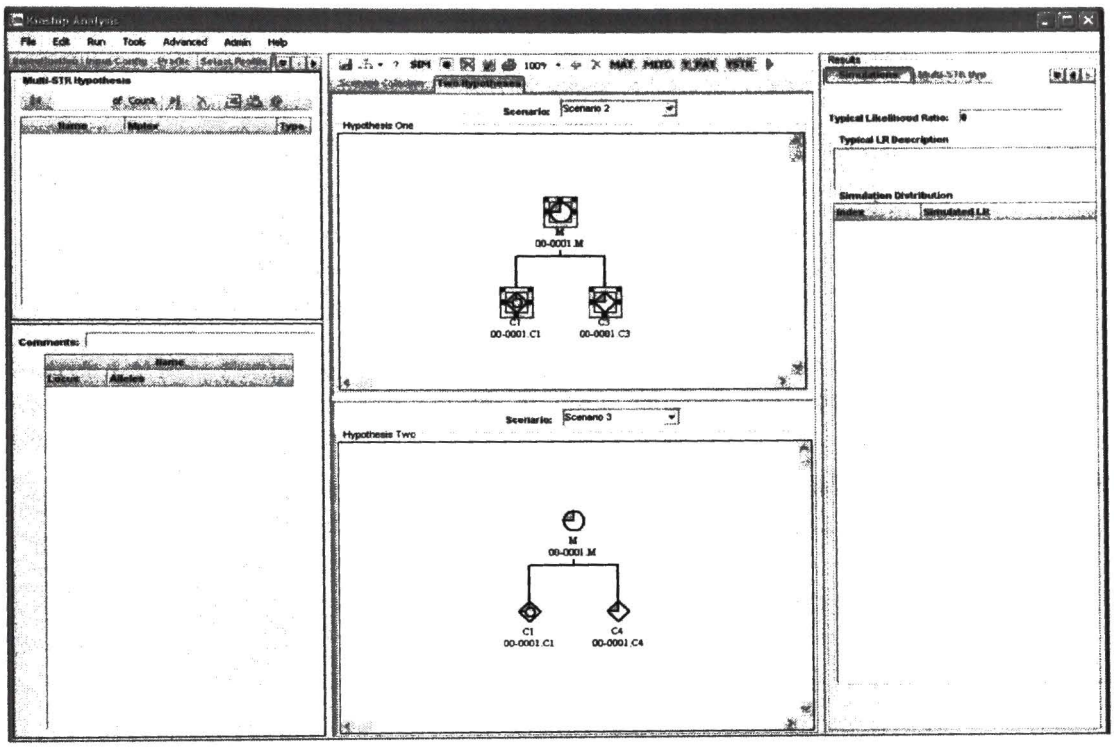


11. Other features with the **Kinship Analysis** window can be found in the **Run** and the **Advanced** tabs on the upper right corner of the window.

12. In the **Run** tab, a UHR search using a particular pedigree can be done when selecting **Against All Unidentified Incident Profiles**.

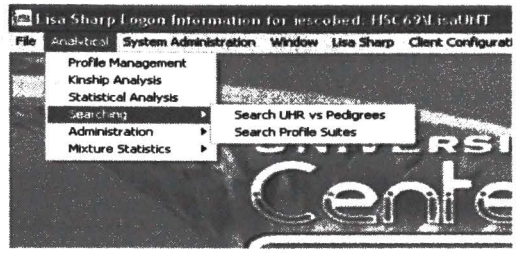


13. Simulation analysis, multi-STR hypothesis analysis, or two hypothesis analysis can be carried out by selecting the options in the **Advanced** tab.

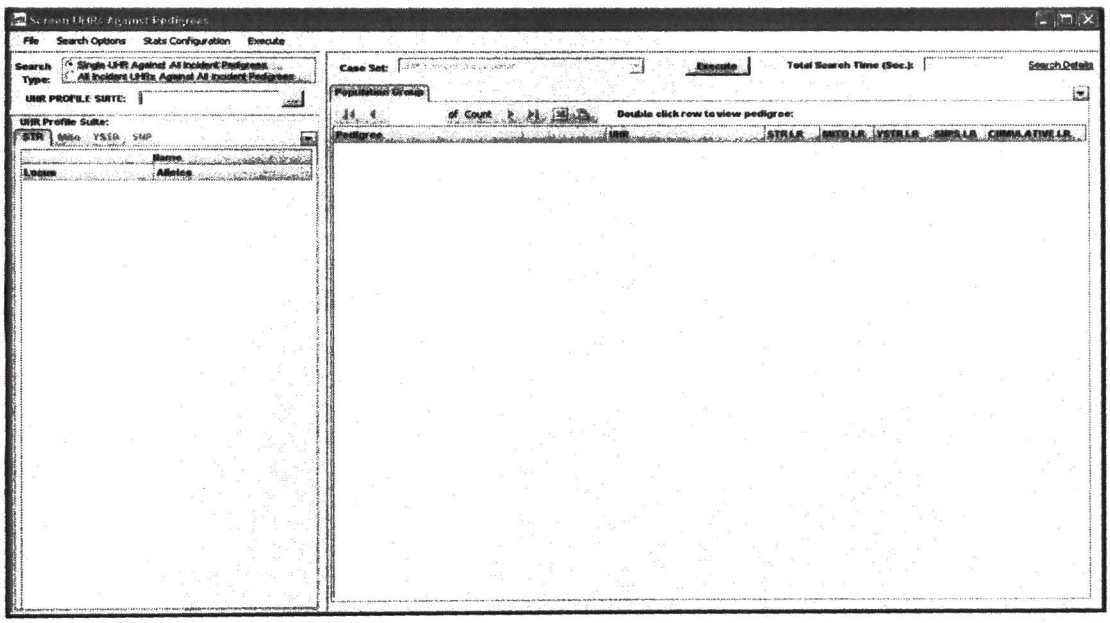


VII. Using the Searching Module for Pedigree Searches

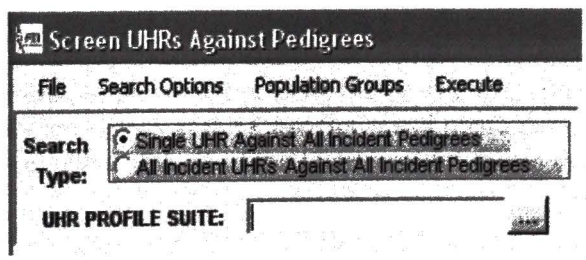
1. Go to the main LISA window, locate **Analytical,** and select **Searching** where you can choose between **Search UHR vs Pedigrees** or **Search Profile Suite**.



2. **Search UHR vs Pedigrees** option is used for searching pedigrees saved in a case set using a selected UHR profile. When this option is selected, the **Screen UHRs Against Pedigrees** window appears.

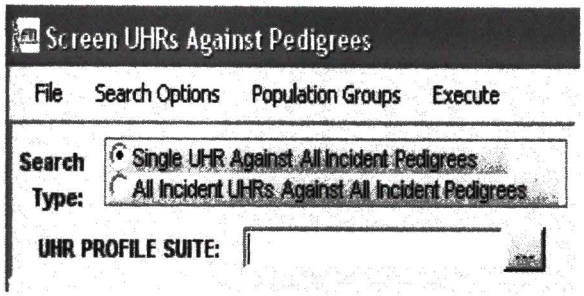


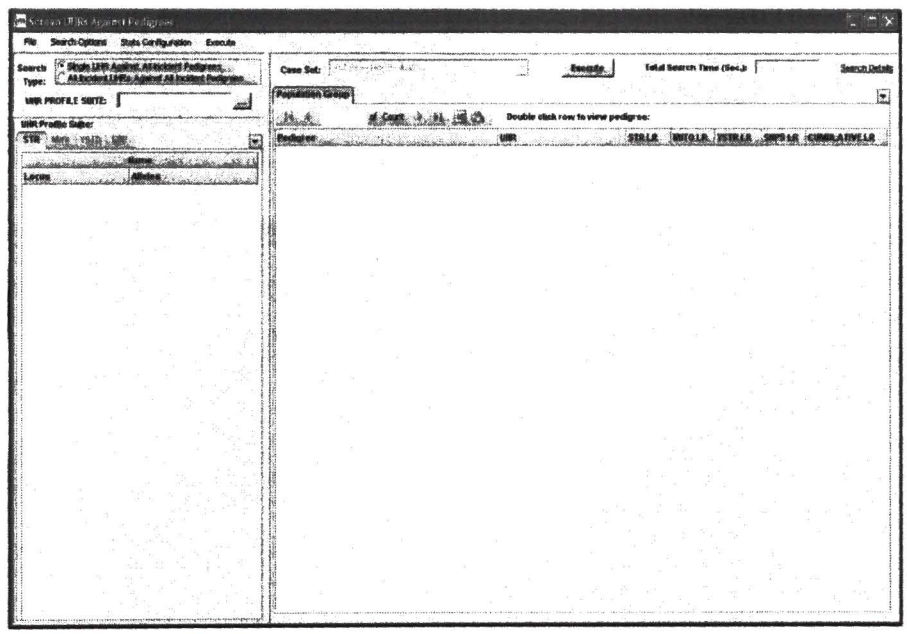
3. To make a search, select the type of search that will be made in the **Search Type** section in the upper left corner of the window.



4. Selecting Single **UHR Against All Incident Pedigrees** allows the user to find an association between a UHR profile and a pedigree by searching through all of the pedigrees found in the default case set.

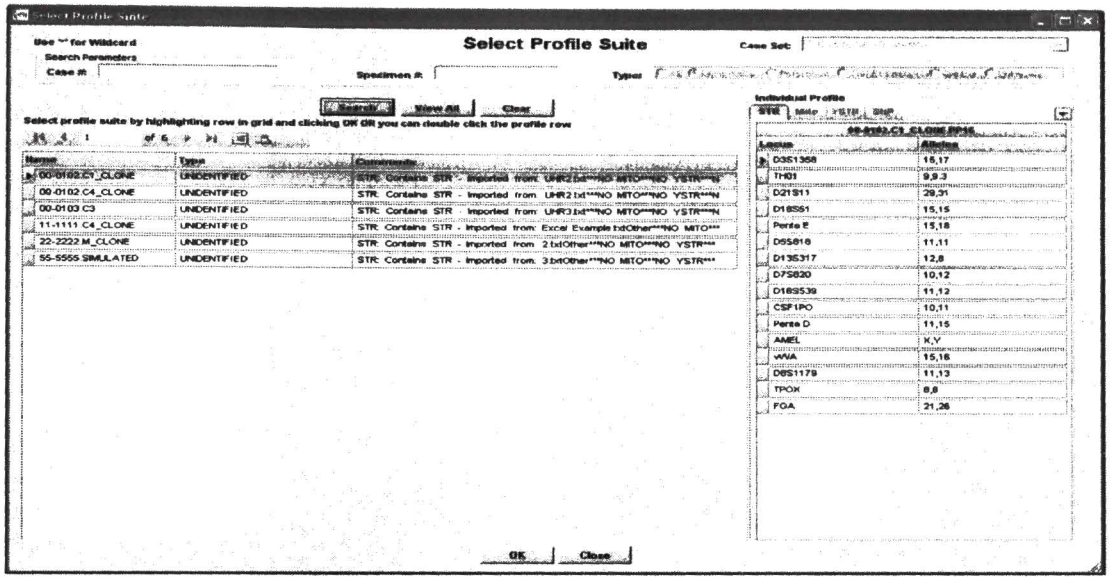
5. Once **Single UHR Against All Incident Pedigrees** is selected, click the tab with 3 black dots next to **UHR PROFILE SUITE** field. Once the icon is clicked, the **UHR Profile Suite** window appears.





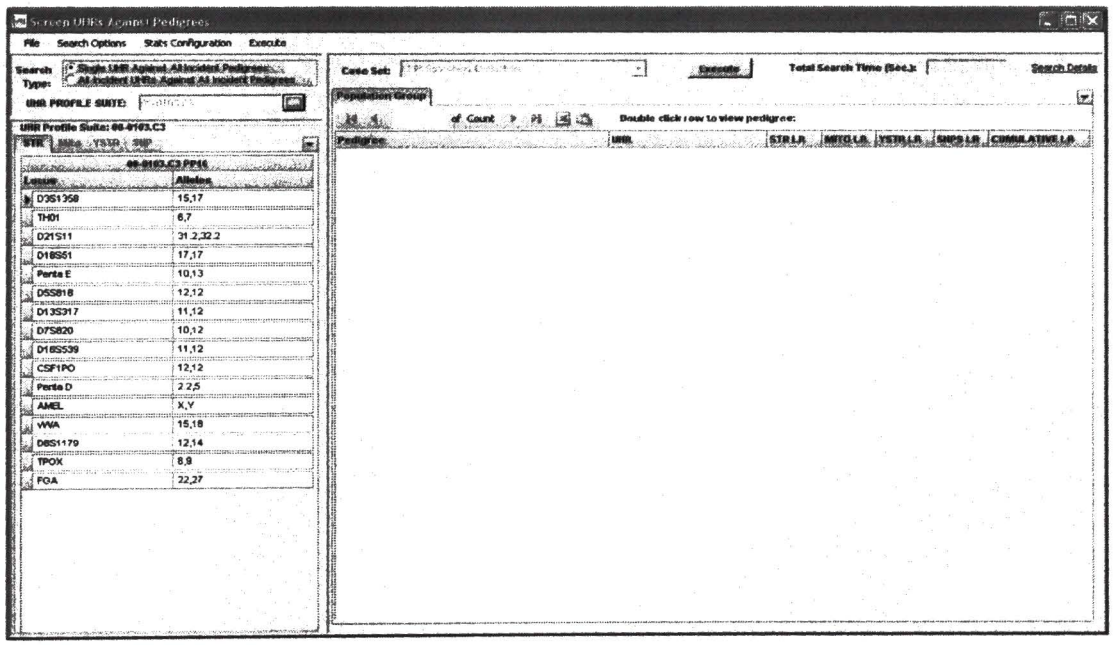
6. In the **Search Profile Suite** window, select UHR the profile from the default case set that will be used for the pedigree search at the **Case Set** scroll menu at the right corner of the window. Click the **Search** tab to list all of the UHR profiles found in the case set. To

narrow down the search, type in the case number or specimen number in the blank spaces in **Search Parameters** section and click the **View All** tab.



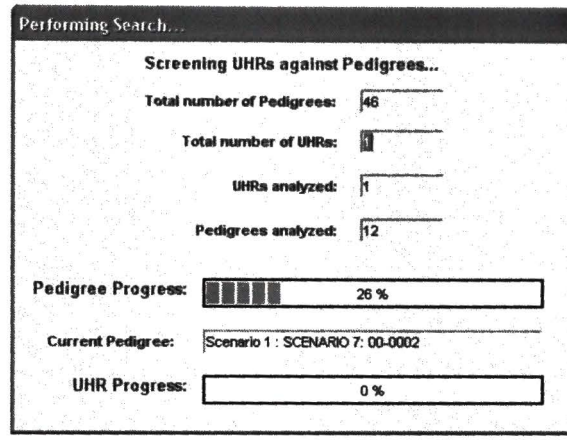
7. Once the results are listed, highlight a particular UHR profile to view the available STR, Mito, Y-STR, and SNP profiles.

8. Click on the highlighted UHR profile in order to be used for the pedigree search. The selected UHR profile (the name and profile) can be seen in the **Screen UHRs Against Pedigrees** window.



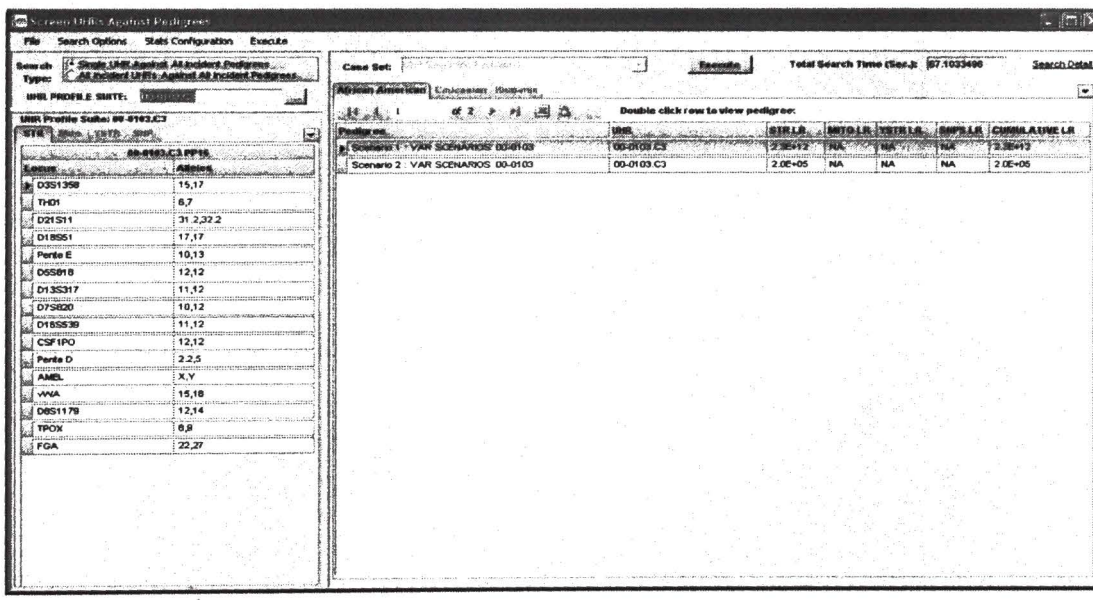
9. After selecting the appropriate UHR profile click the **Execute** tab on top of the window to run the search. To change the default case set go to **Lab Configuration** tab at the main LISA window.

10. A **Performing Search** window will appear indicating the number of pedigrees and UHR profiles being analyzed, and the progress of the search.

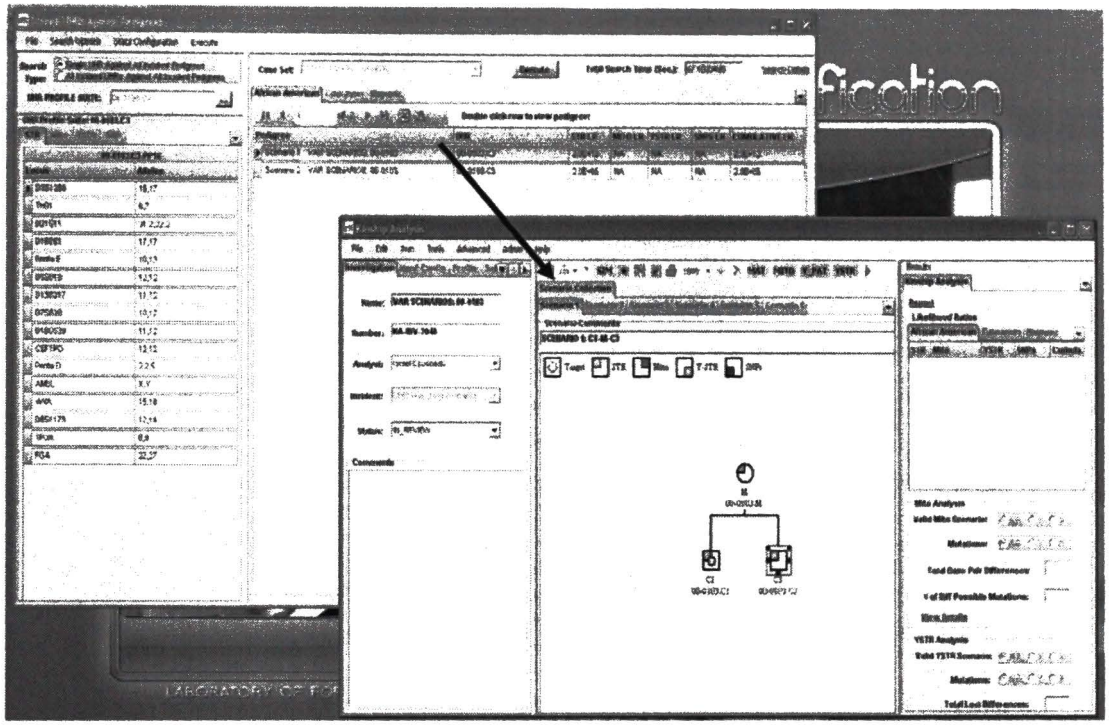


11. Once the search is finished, the total time that it took for the search to be performed will be shown in the **Total Search Time (Sec)** box.

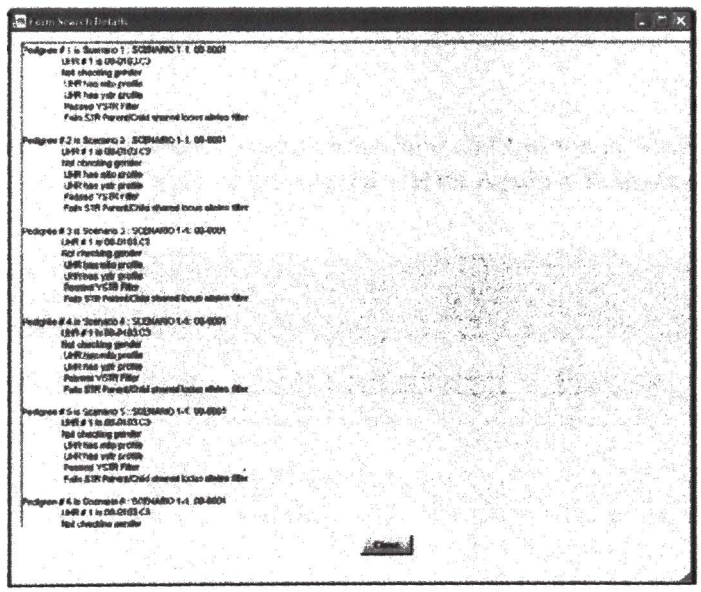
12. A list of pedigrees or scenarios where associations were made with the UHR profile is displayed in each of the three major population group tabs. The associations are ranked according to the cumulative highest likelihood ratio. The likelihood ratios for STR, Mito, Y-STR and SNP data are also shown.



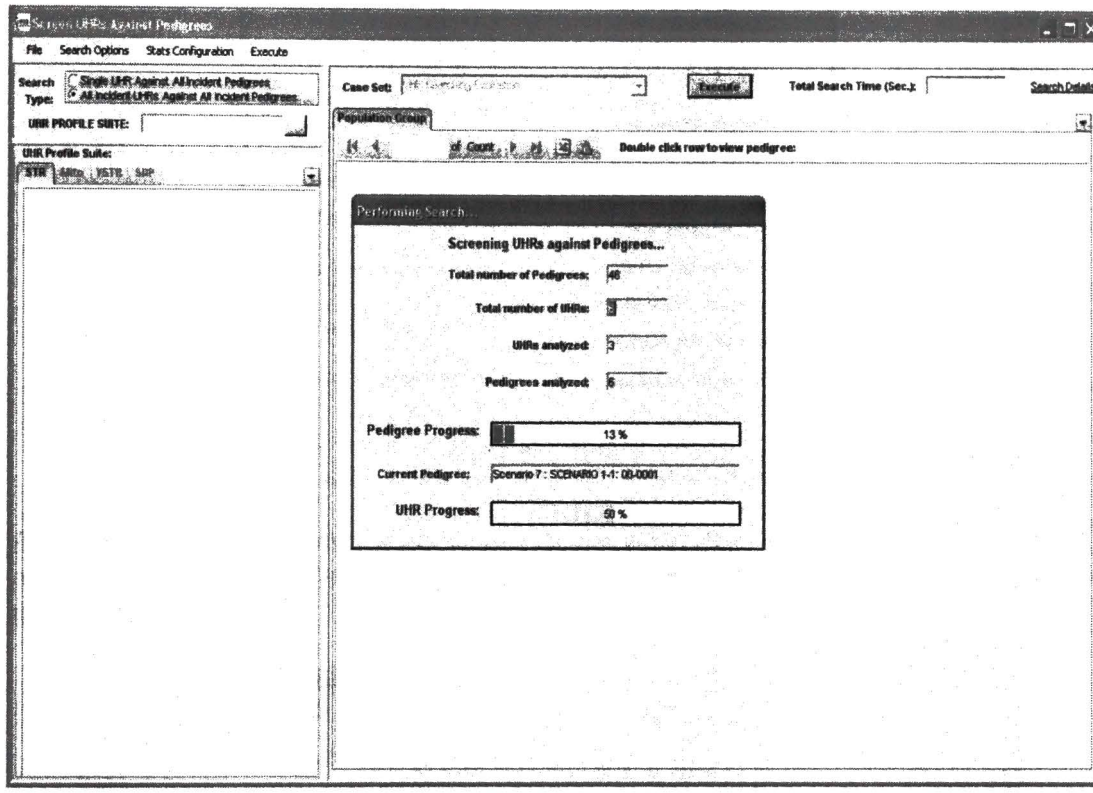
13. Double click any result to view the contents of the listed investigation file, and to find the pedigree that made the association with UHR profile based on the listed scenario number.



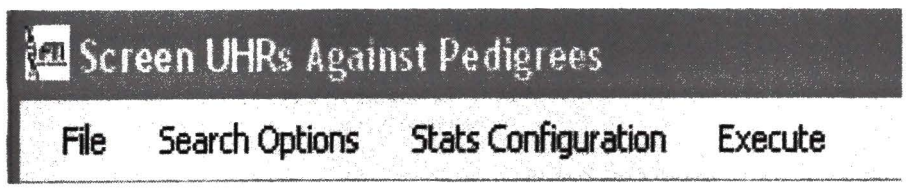
14. To look at the details of the search, click the blue **Search Details** link at the top right corner of the **Screen UHRs Against Pedigrees** window.



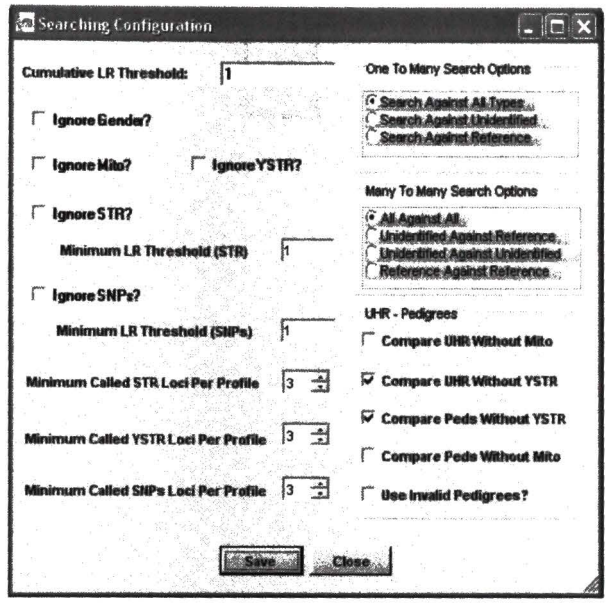
15. To search all UHR profiles from particular case set against all pedigrees in particular case set can be done by selecting **All Incident UHRs Against All Incident Pedigrees** in the **Search Type** section. Once this option is selected, click the **Execute** tab. The Performing Search window will appear to show the status of the search, and the results will listed based on the highest cumulative LR.



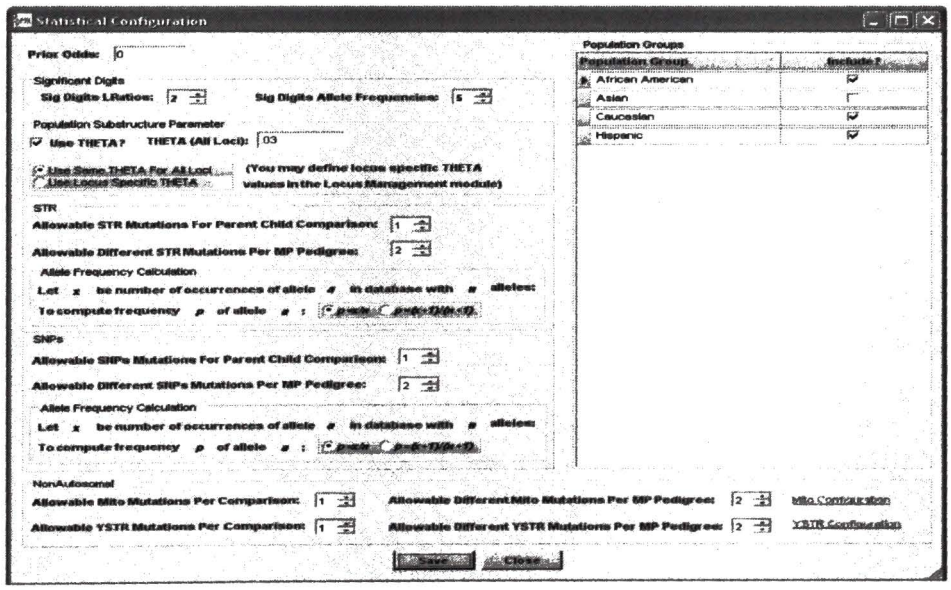
16. To change the settings of any search, select either **Search Options** or **Stats Configuration** tabs at the top right of the **Screen UHRs Against Pedigrees** window.



17. A **Searching Configuration** window will appear when **Search Options** is selected where a LR threshold can be set, modify search parameters, or shutdown Mito and Y-STR filters.



18. A **Statistical Configuration** window will appear when **Stats Configuration**, where population database can be selected, assigns prior odds/ theta, and configure other search parameters.



19. To save or print the results of the search, click the excel icon in **Screen UHRs Against Pedigrees** window to save the results in a excel spreadsheet

REFERENCES

1. Aryes, K.L. Paternity Index calculation when some individuals share common ancestry. Forensic Science International 2005; 151: 101-103.

2. Brenner, C.H. What is DNA•VIEW™? http://dna-view.com/dnaview.html.

3. Brenner, C.H. Kinship by Hand. 2004. http://courses.hsc.unt.edu/gsbs/ForensicGen/GDA/GDA_PDF/Kinship%20by%20hand.pdf

4. Brenner, C.H. Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. Forensic Science International 2006; 157: 172-180.

5. Brenner, C.H. Symbolic Kinship Program. Genetics 1997; 145: 535-542.

6. Brenner, C.H., and Weir, B.S. Issues and strategies in the DNA identification of World Trade Center victims. Theoretical Population Biology 2003; 63: 173-178.

7. Birus, I., et al. How High Should Paternity Index Be for Reliable Identification of War Victims by DNA Typing? Croatian Medical Journal 2003; 44(3): 322-326.

8. Buckleton, J. Triggs, C.M., and Walsh, S.J., editors. Forensic DNA Evidence Interpretation. Boca Raton: CRC Press, 2005.

9. Budowle, B., Bieber, F.R., and Eisenberg, A.J. Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification. Legal Medicine 2005; 7: 230-243.

10. Dajda, T., and Jung, M. LR-calculation of any kinship situation using a new graphical interface: Generate two or more hypotheses, draw the family trees and assign the DNA profiles to person symbols. International Congress Series 2006; 1288: 474-476.

11. Egeland, T., et al. Beyond traditional paternity and identification cases: Selecting the most probable pedigree. Forensic Science International 2000; 110: 47-59.

12. Egeland, T., and Mostad, P.F. Statistical Genetics and Genetical Statistics: a Forensic Perspective. Scandinavian Journal of Statistics 2002; 29: 297-307.

13. Future Technologies Inc. FTI: Aiding DNA Identification of September 11 Attack Victims. 2005. http://www.ftechi.com/newsroom.html.

14. Fung, W.K. User-friendly programs for easy calculations in paternity testing and kinship determinations. Forensic Science International 2003; 136: 22-34.

15. Fung, W.K., Carracedo, A., and Hu, Y.G. Testing for kinship in a subdivided population. Forensic Science International 2003; 135: 105-109.

16. Gjertson, D.W., et al. ISFG: Recommendations on biostatistics in paternity testing. Forensic Science International: Genetics 2007; 1: 223-231.

17. Irwin, J.A., et al. DNA Identification of "Earthquake McGoon" 50 Years Postmortem. Journal of Forensic Science 2007; 52(5): 1115-1118.

18. Leclair, B. Large-scale comparative genotyping and kinship analysis: evolution in its use for human identification in mass fatality incidents and missing persons databasing. International Congress Series 2004; 1261: 42-44.

19. Leclair, B., et al. Bioinformatics and Human Identification in Mass Fatality Incidents: The World Trade Center Disaster. Journal of Forensic Science 2007; 52(4): 806-819.

20. Leclair, B., et al. Enhanced Kinship Analysis and STR-based DNA Typing for Human Identification in Mass Fatality Incidents: The Swissair Flight 111 Disaster. Journal of Forensic Science 2007; 52(4): 806-819.

21. Li, C.C., and Sacks, L. The Derivation of Joint Distribution and Correlation between Relatives by the Use of Stochastic Matrices. Biometrics 1954; 10(3): 347-360.

22. Myers, S. DOJ View Version 2.0, 3-20-07. Available from Cal DOJ Jan Bashinski DNA Lab.

23. National Institute of Justice. Lessons Learned From 9/11: DNA Identification in Mass Fatality Incidents. U.S. Department of Justice and National Institute of Justice 2006; 1-143. http://www.dna.gov

24. Poetsch, M., et al. The problem of single parent/child paternity analysis-Practical results involving 336 children and 348 unrelated men. Forensic Science International 2006; 159: 98-103.

25. Presciuttini, S., et al. Allele sharing in first-degree and unrelated pairs of individuals in the Ge.F.I AmpFISTR® Profiler Plus™ database. Forensic Science International 2003; 131: 85-89.

26. Prinz, M., et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for

disaster victim identification (DVI). Forensic Science International: Genetics 2007; 1: 3-12.

27. Riancho, J.A., and Zarrabeitia, M.T. A Windows-based software for Common paternity and sibling analyses. Forensic Science International 2003; 135: 232-234.

28. Tzeng, C.H., et al. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan. Transfusion 2000; 40:840-845.

29. Wenk, R.E., and Chiafari, F.A. Distinguishing full siblings from half-siblings in limited pedigrees. Transfusion 2000; 40:44-47.

30. Wurmb-Schwark, N.V., et al. Possible pitfalls in motherless paternity analysis with related putative fathers. Forensic Science International 2006; 159: 92-97.

31. Future Technologies Inc. Laboratory Information Systems Applications (LISA): Beta Version. Accessed June 2008.

32. DNA Advisory Board. Quality Assurance Standards for Forensic DNA testing Laboratories. Forensic Science Communications 2000; 2(3): 1-17.

33. Future Technologies Inc. DNA Biometrics and DNA Laboratory Information Management And Analysis System Development. http://www.ftechi.com/dna_biometrics.html.

34. SoftGenetics®. GeneMarker® HID: STR Human Identity Software. 2008. http://www.softgenetics.com/GeneMarkerHID.html