

DNA Mixture Analysis:

Principles and Practice of Mixture Interpretation and Statistical Analysis
Using the SWGDAM STR Interpretation Guidelines

Fundamentals of Interpreting STR Mixtures

Michael Adamowicz

University of New Haven

AAFS 2011 Workshop #17

Chicago, IL

February 22, 2011



NIST



Why interpret mixtures at all?

If little or no attempt is made to interpret the data contained in the mixture profile, then you are not using the evidence to its full potential!

But how do we sort out all of the variables and choices?

Use the SWGDAM guidelines and the literature with your own validation studies to create a documented mixture interpretation **process** for all of the analysts in your laboratory

A major challenge in mixture interpretation has been achieving consistency among analysts within the same laboratory;

Same data, different answers → **Not a Good Situation**

The solution is to create a clearly defined and documented process for all analysts to follow which will provide the tools to consistently interpret mixtures

It is impossible to cover every contingency, but a well defined mixture interpretation process should allow for the consistent analysis of all data

What does this process look like?

It can take any number of forms

- Flow Chart
- Written Instructions
- Computer Macro/Decision Tree

The format just needs to be clear so analysts can employ it effectively

The process really is the **BIG PICTURE:**

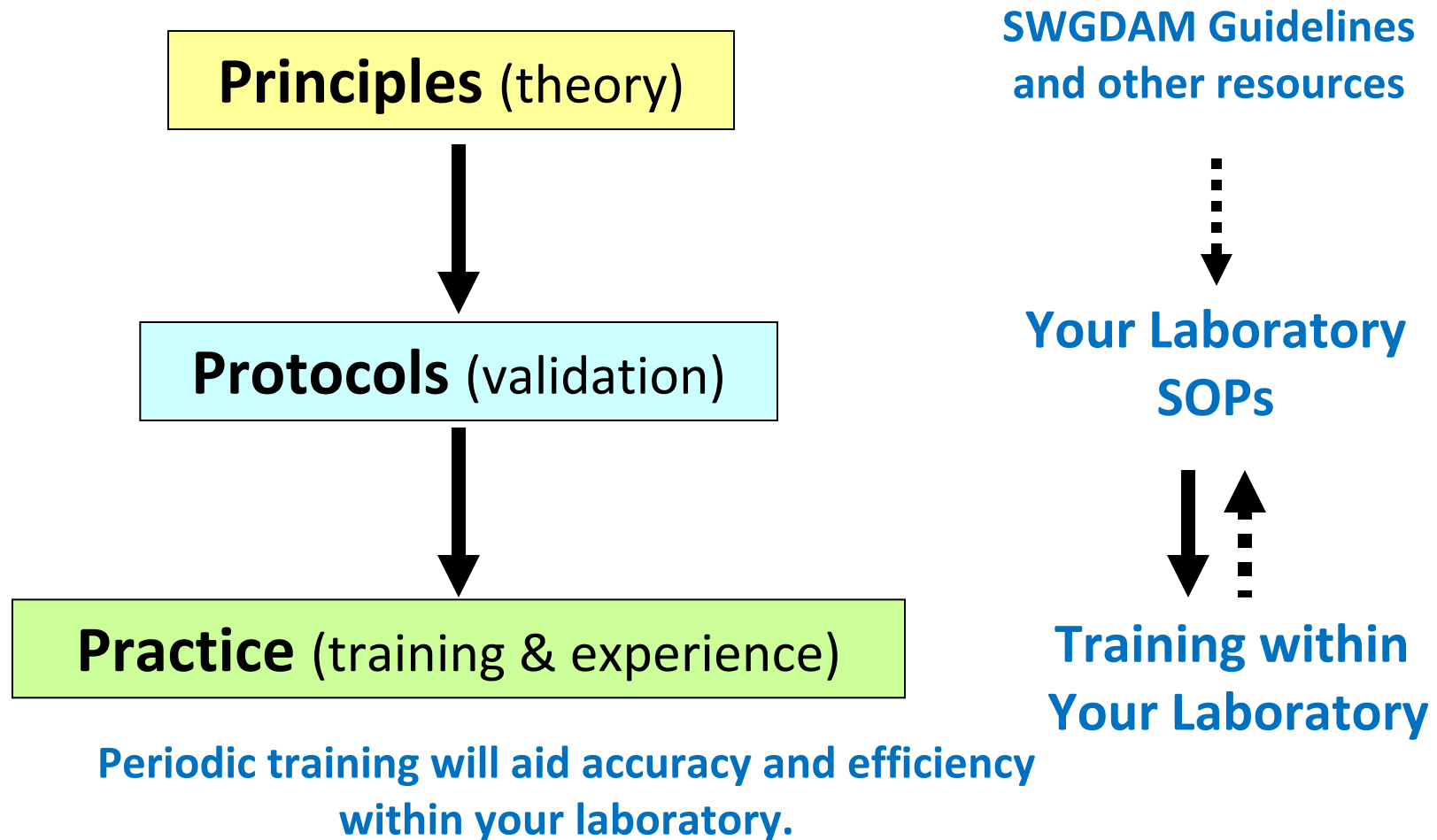
STR Data  Interpretive statement

Benefits of a documented Mixture Interpretation Process include;

- Increased analyst confidence
- Less time in technical review
- Documentation for discovery
- Consistency in reports and testimony

What should be included in a mixture interpretation process

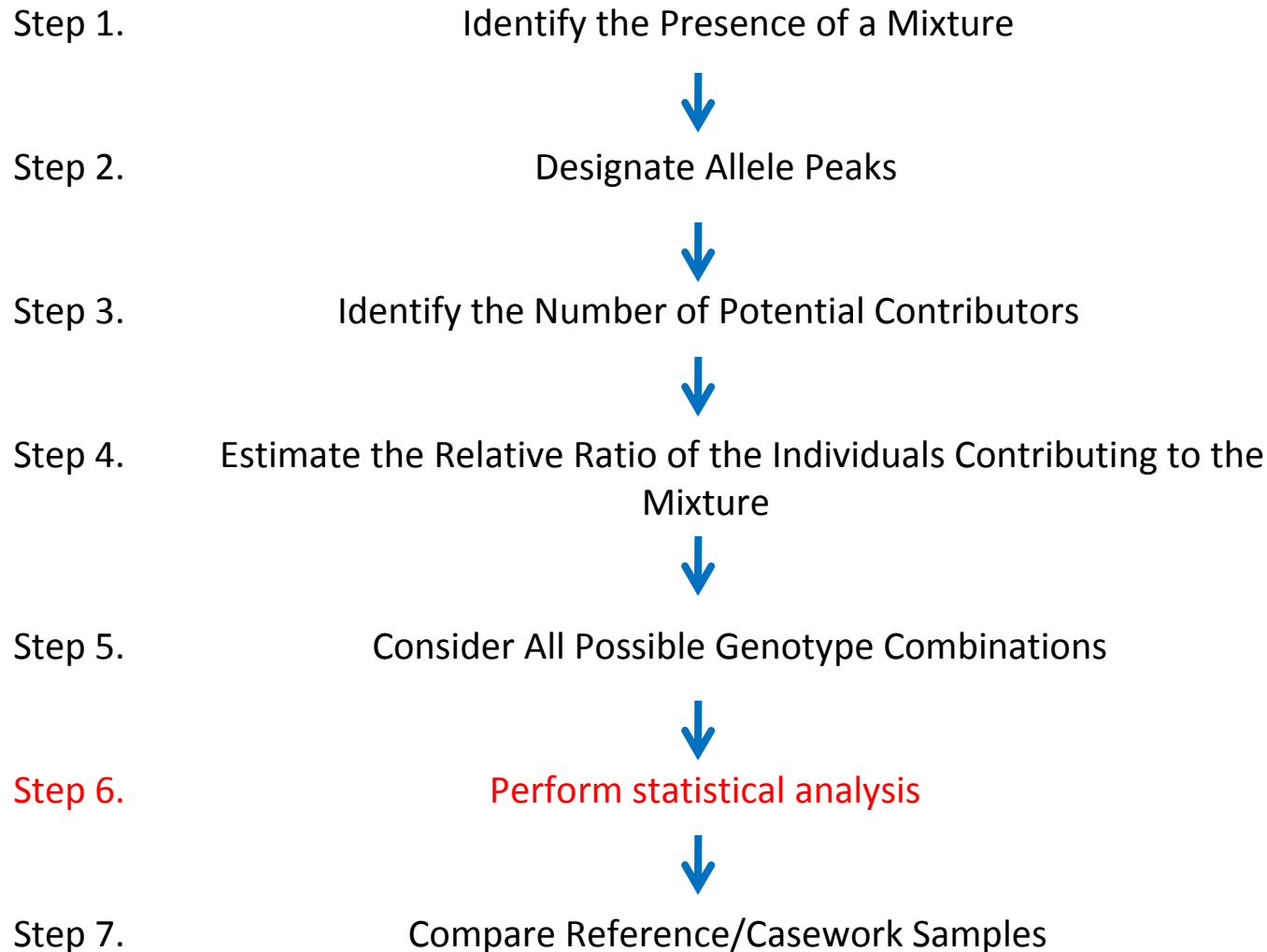
Elements of DNA Mixture Interpretation



Consistency amongst analysts

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Assumptions

Assumptions are an integral part of the process of mixture interpretation

Common assumptions;

- Possible Genotype Combinations
- # of contributors
- Intimate samples (i.e. the assumed contributor used to deduce obligate alleles)

Assumptions are not guesses, they are based on validation and experience

3.5.2. The laboratory should define and document what, if any, assumptions are used in a particular mixture deconvolution.

3.6.5. Because assumptions regarding the origin of evidence or the number of contributors to a mixture can impact comparisons, the laboratory should establish guidelines for documenting any assumptions that are made when formulating conclusions.

Validation is Critical

Mixture validation will define

- Thresholds
- Peak height ratios (PHRs)
- Stutter
- Stochastic behavior



Tools

All of these will provide the framework for your assumptions and will help to define your protocols and your overall **process**

Analytical Threshold

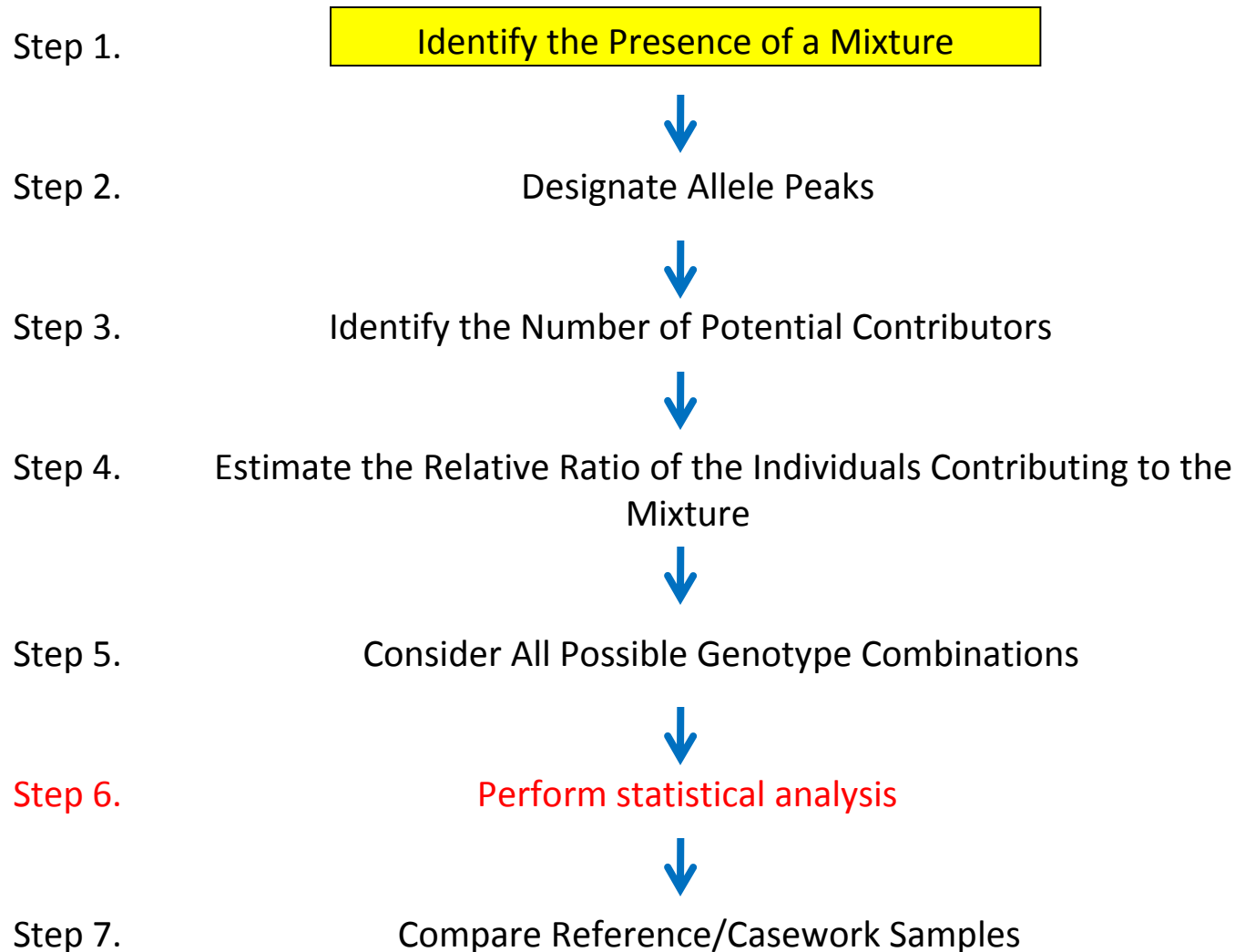
- An analytical threshold must be established that operationally defines the minimum peak height in RFUs for confidently ascribing a true PCR amplicon peak
- Defines when confidence is high for peak assignment
- The analytical threshold is based on signal to noise ratio (often rounded up to a convenient number such as 50 RFUs)

Stochastic Threshold

- This threshold is defined as the value above which it is reasonable to assume that allelic dropout has not occurred within a single-source sample
- It is the minimum peak height in RFUs that all amplicon peaks in a locus must display to conclude with confidence that no genetic components of the sample have failed to be detected due to stochastic effects (*low copy template*)
- Critical in identifying homozygotes
- ALLELE DROP OUT
- This threshold may become very important when calculating statistics on mixture samples

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Identify the Presence of a Mixture

What constitutes a mixture?

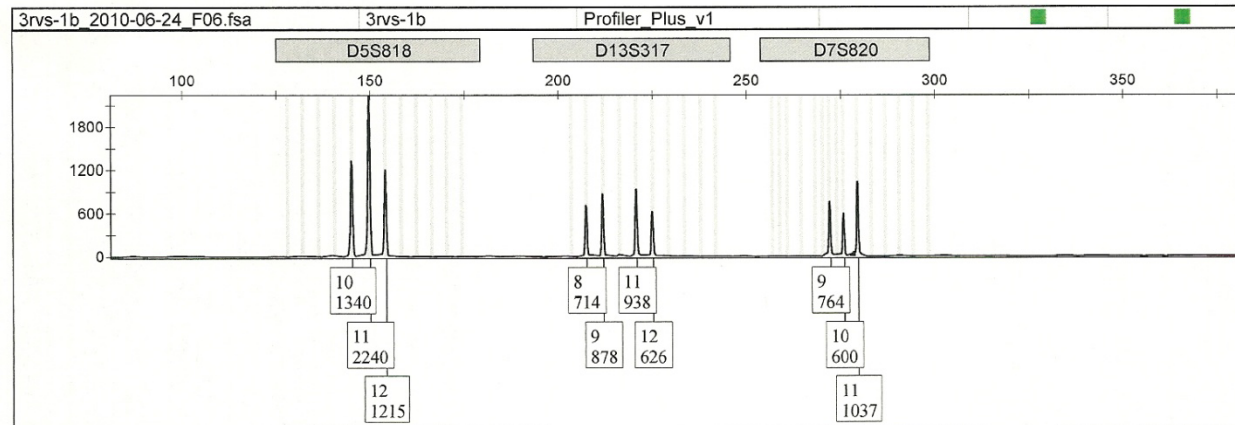
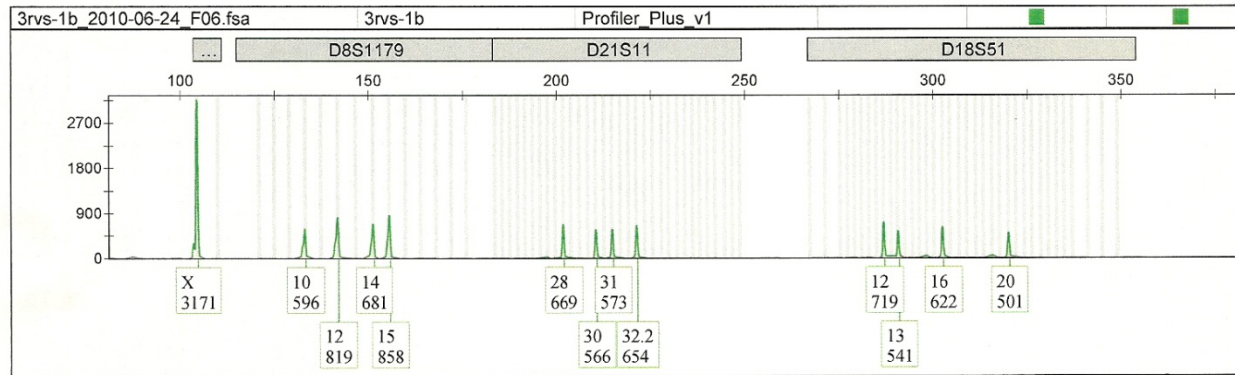
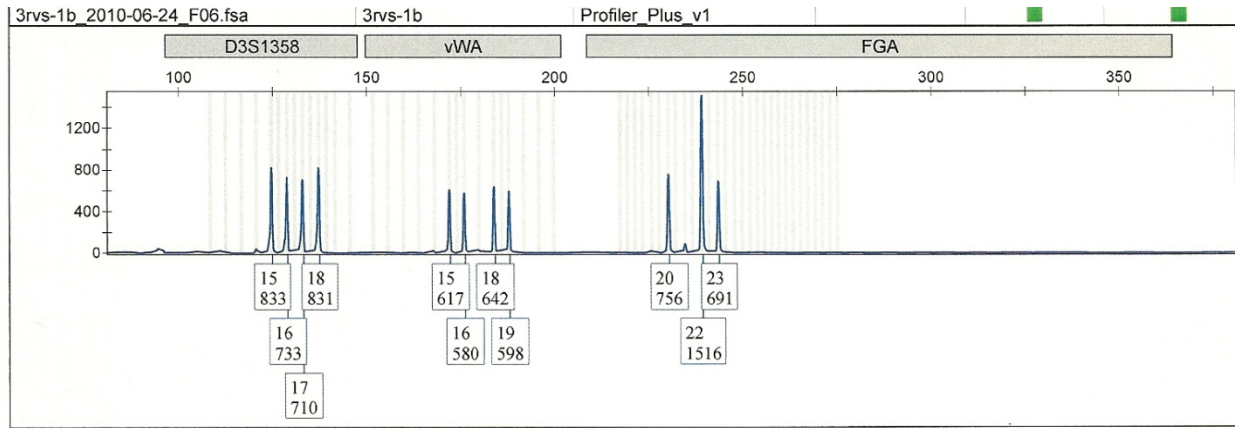
- More than 2 alleles at a single locus (note exception of tri-allelic patterns)
- More than 2 alleles at one or more loci (tri-allelic pattern as exception)
- Three or more alleles at two or more loci or 4 or more alleles detected at one locus
- The presence of peak height imbalance at 2 or more loci may be due to the presence of a mixture, particularly if the peak heights of the alleles are not in the stochastic range

There are multiple, valid options

Be clear on which is used in your process and document the results

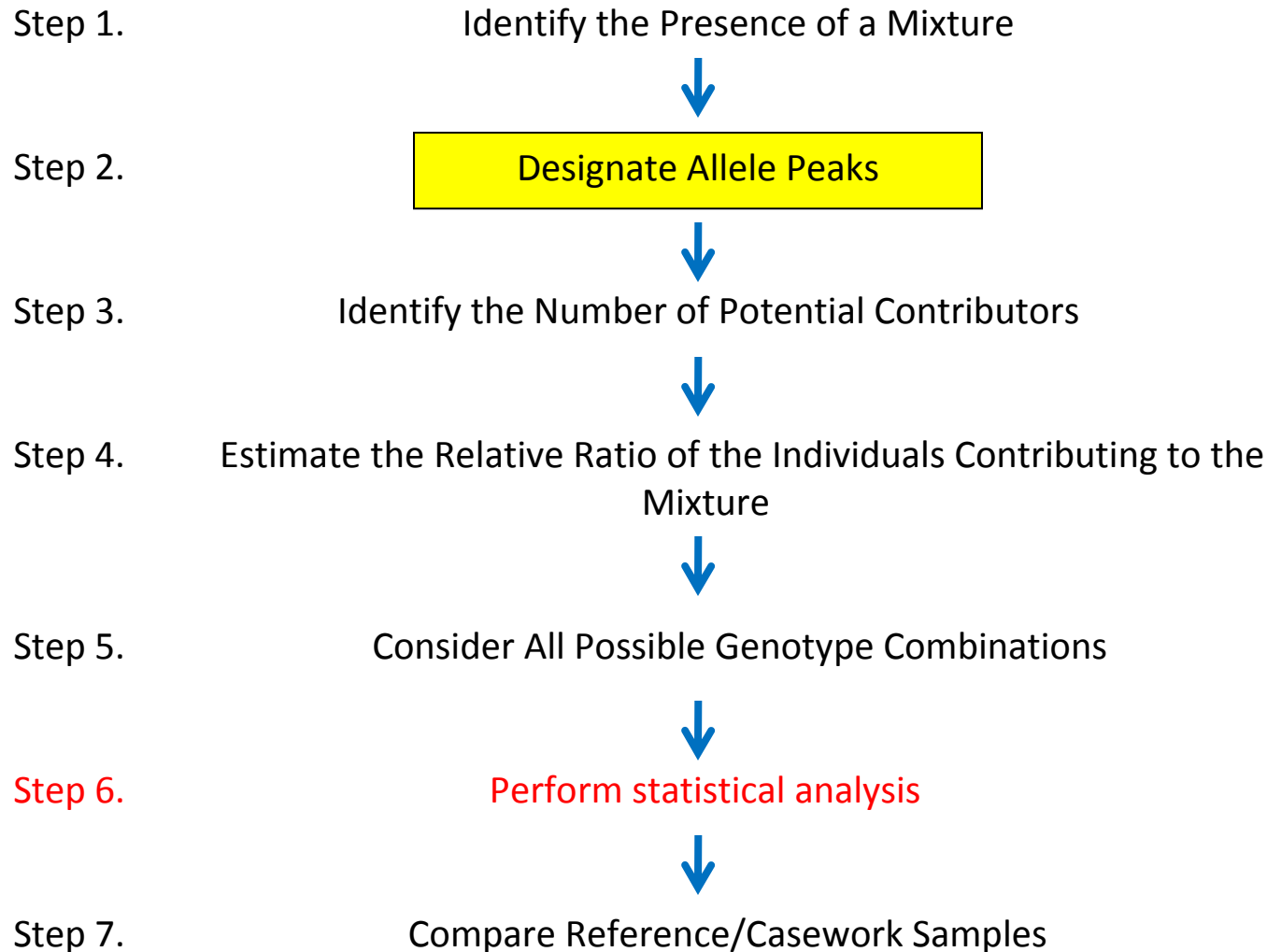
The entire profile needs to be considered when declaring a mixture, **use all available data**

Guideline **3.4** Number of Contributors to a DNA Profile



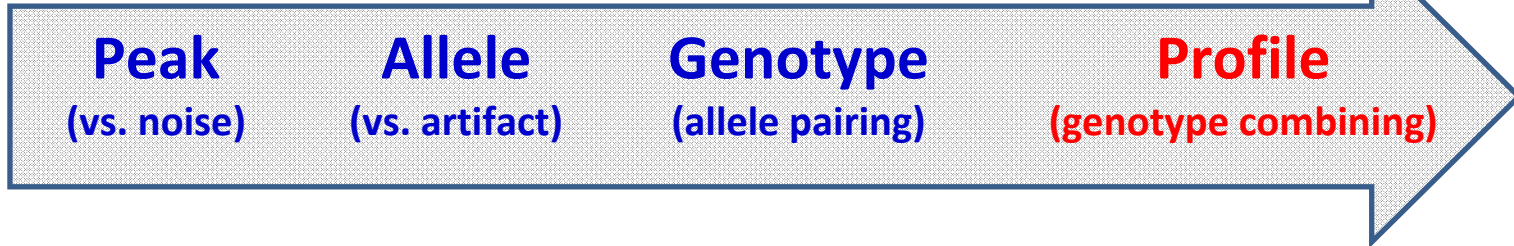
Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]

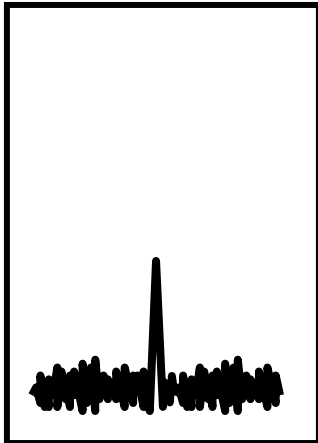


Mixture Interpretation Protocols Build on Single-Source Sample Information

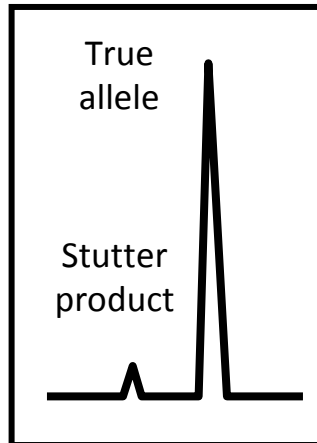
The Steps of Data Interpretation



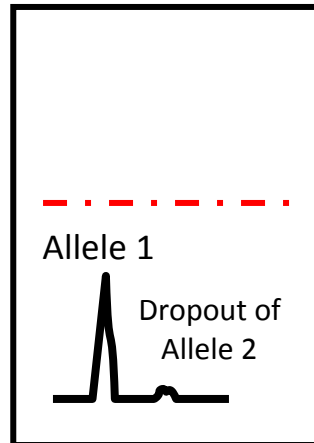
**Analytical
Threshold**



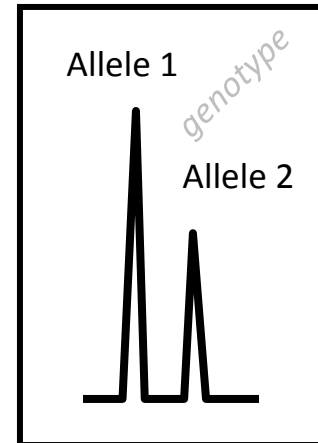
**Expected
Stutter %**



**Stochastic
Threshold**



**Peak Height
Ratio (PHR)**



Moving from individual locus genotypes to profiles of potential contributors to the mixture is dependent on mixture ratios and numbers of contributors

Complications in Mixture Identification

Tri-allelic patterns and primer binding site mutations

- Very rare to have at more than 1 locus

Stutter

Again, multiple options exist

- Guidelines 3.5.8.1, 3.5.8.2, and 3.5.8.3 apply
- Hard cut-off, never override stutter percentage cut-offs regardless of mixture
- Case-to-case; Calling or eliminating a sub-threshold stutter position peak based on how it fits with the rest of the mixture profile (i.e. consistent peak height, does or does not fit with other alleles)

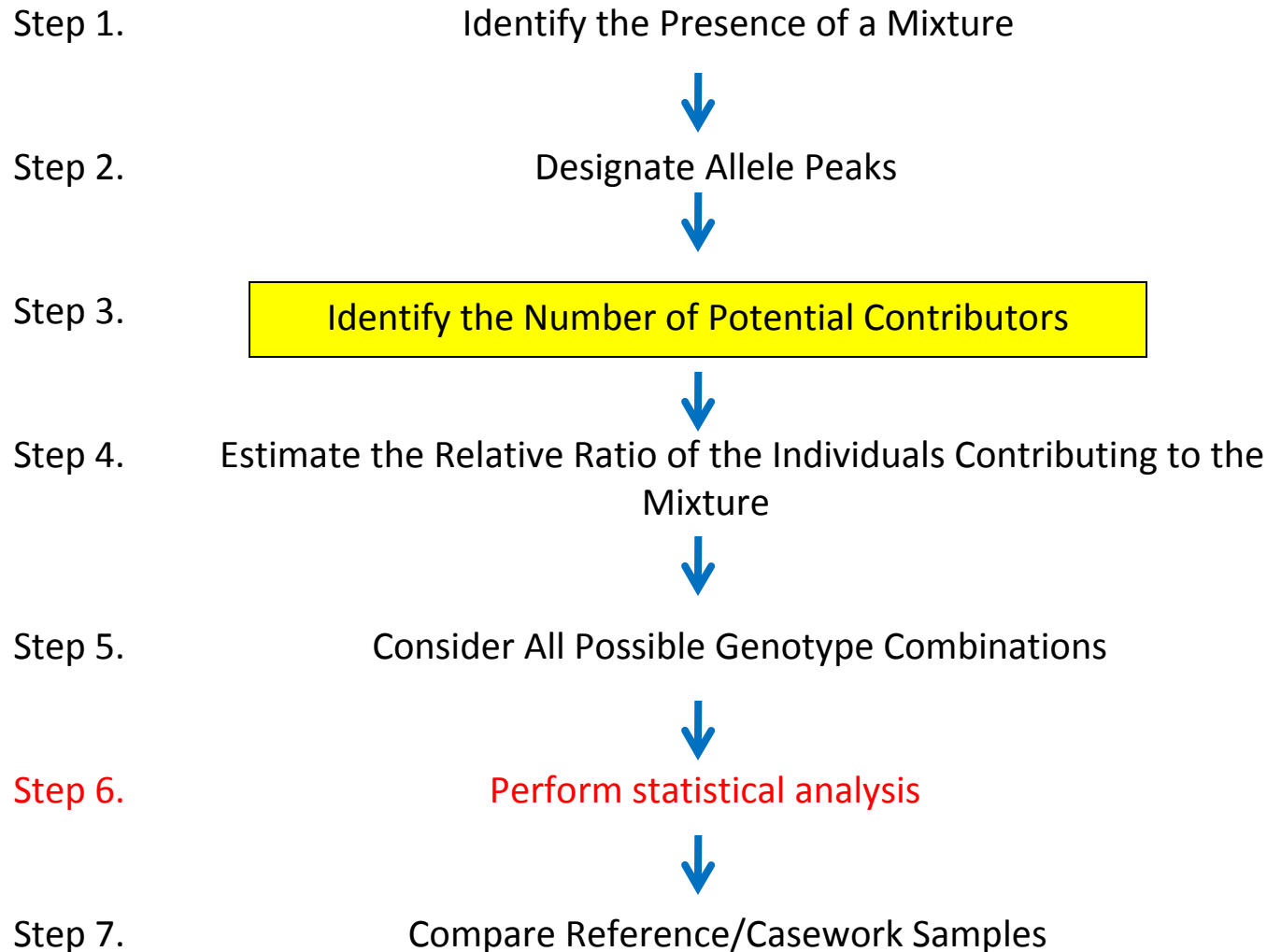
The SOP must clearly define criteria for

- when a peak which is at-or-below the stutter threshold will be identified as a mixture peak and not stutter
- if it is identified as a peak, how this will be documented
- whether or not this locus will be used for statistics

Otherwise mixture interpretation may become gray

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Identify the Number of Potential Contributors

Usually based on the largest number of identifiable alleles at one or more loci using the assumption that there are 2 alleles per locus per individual

Expected results when estimating the number of contributors

- If there are 4 or fewer alleles observed at every locus across a profile, then 2 contributors are most likely
- If there are a maximum of 5 or 6 alleles at any locus, then 3 contributors are likely
- If there are more than 6 alleles present in a single locus, then 4 or more contributors are likely

These results are for unrelated people

Identify the Number of Potential Contributors

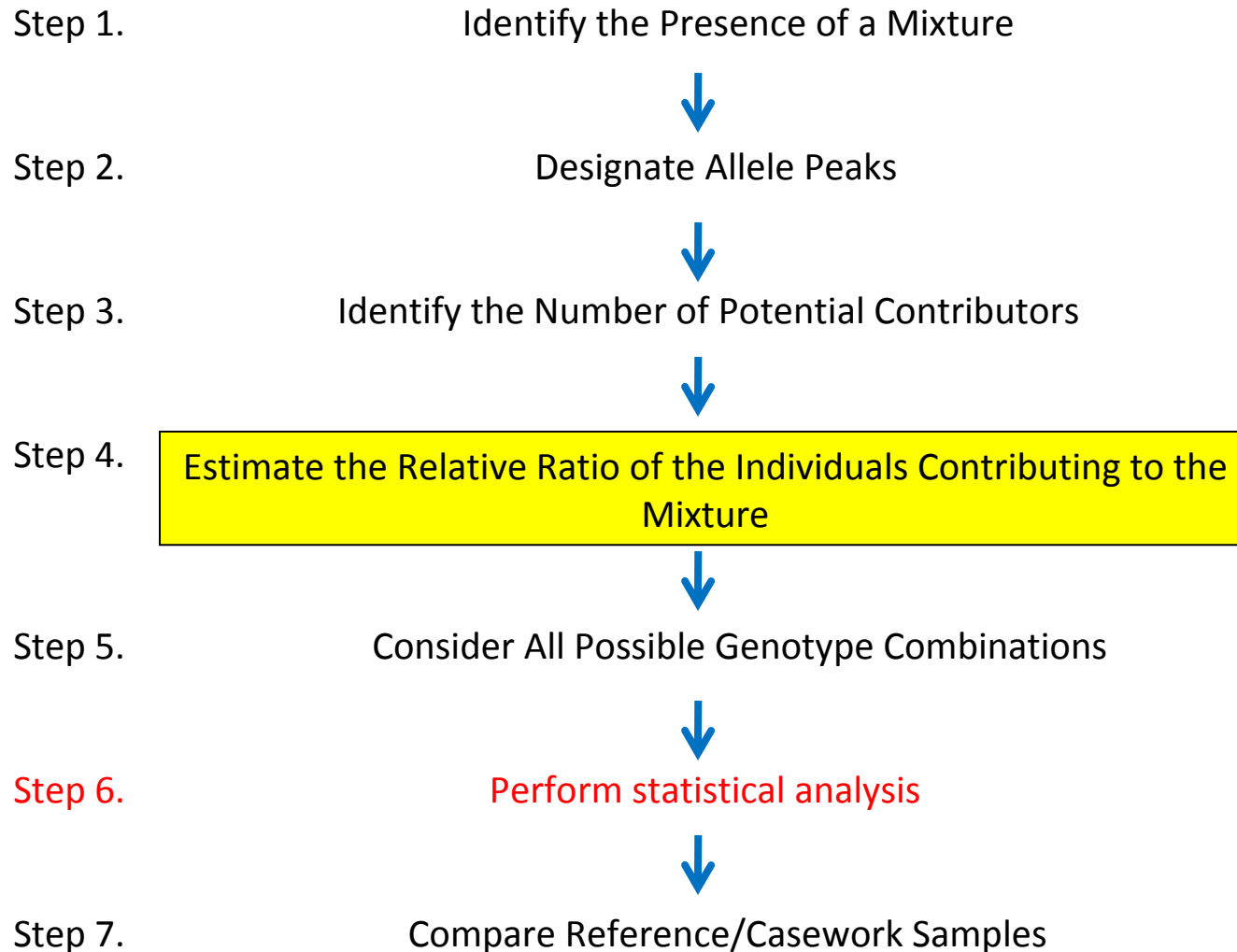
Commonly expressed as a lower bound;

“Five alleles were observed at locus D3S1358, therefore this profile is a mixture of at least three individuals.”

3.5.2.2. If assumptions are made as to the number of contributors, additional information such as the number of alleles at a given locus and the relative peak heights can be used to distinguish major and minor contributors.

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Peak Height Ratios (PHRs)

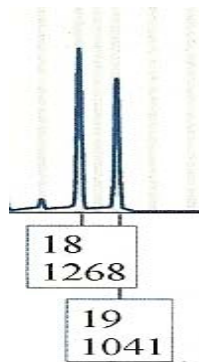
To use PHRs in mixture interpretation, validation data is needed to assess what is normal PHR variation in heterozygous sister alleles

Options

- Same PHR for all loci
- PHR determined for each locus independently
- A series of locus specific values across multiple peak height ranges
- Common PHRs range from 60 – 70%

Guideline 3.3 (3.3.2. PHR requirements are only applicable to allelic peaks that meet or exceed the stochastic threshold.)

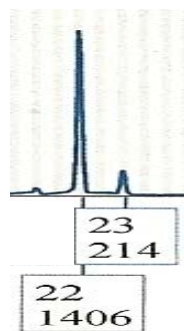
Peak Height Ratios



At every locus

Lower rfu Peak

Higher rfu Peak



$$\frac{1041}{1268} = 82\% \rightarrow \text{Could be sister alleles}$$

$$\frac{214}{1406} = 15\% \rightarrow \text{Could be sister alleles}$$

Exceptions

- primer binding site mutations
- inhibition
- degraded DNA
- stochastic range


Other loci will aid in determining if any of the exceptions are present

Estimate the Relative Ratio of the Individuals Contributing to the Mixture

Based on PHRs, some mixture profiles may be able to be **deconvoluted** or separated into Major and Minor contributors by calculating a Mixture Ratio

The Mixture Ratio used for distinguishing between major and minor contributors should be documented and a concise part of the mixture interpretation process employed

Mixture Ratios of 3:1 or 4:1 have been reported

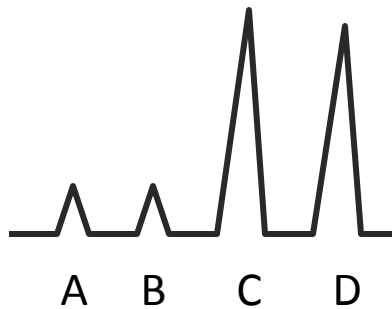
Not every locus may be resolved  **Allele Sharing**

Guidelines **3.5.1, 3.5.2, 3.5.3, 3.5.4, 3.5.5, 3.5.6**

3.5.2. The laboratory should define and document what, if any, assumptions are used in a particular mixture deconvolution.

Calculate the Major/Minor Ratio

Use loci with the maximum # of alleles



Possible allelic combinations are AB, AC, AD, BC, BD, and CD **assuming** 2 contributors

But which combinations are most likely?

Determine using PHRs of potential sister alleles

AB = 90% AC ~~= 32%~~ AD ~~= 33%~~ BC ~~= 30%~~ BD ~~= 31%~~ CD = 85%

Best combination AB and CD \longrightarrow calculate mixture ratio

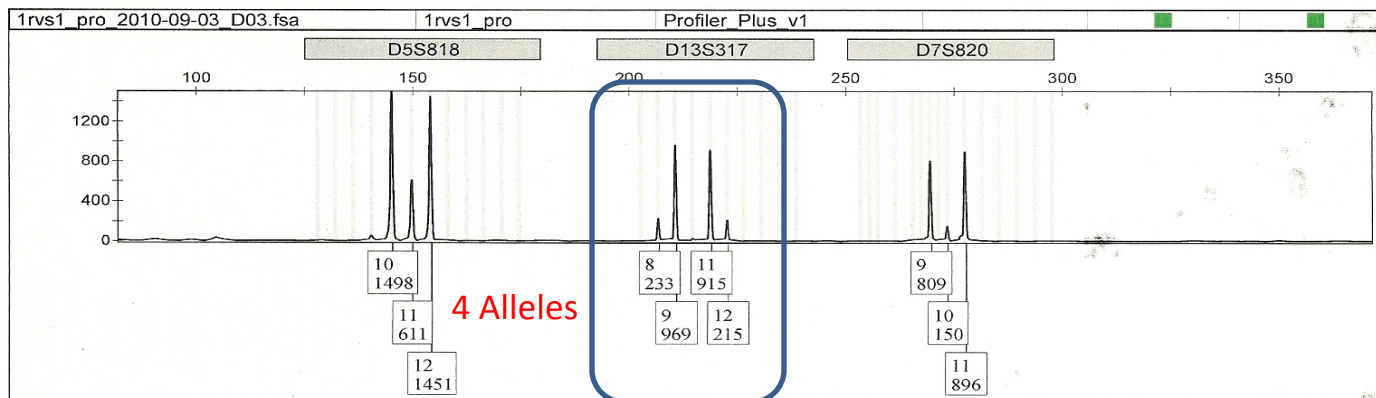
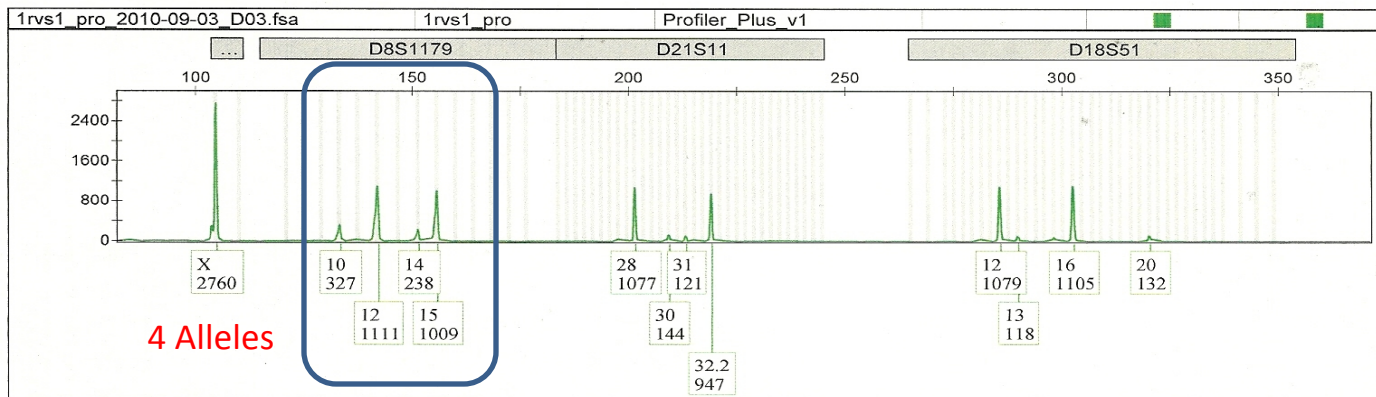
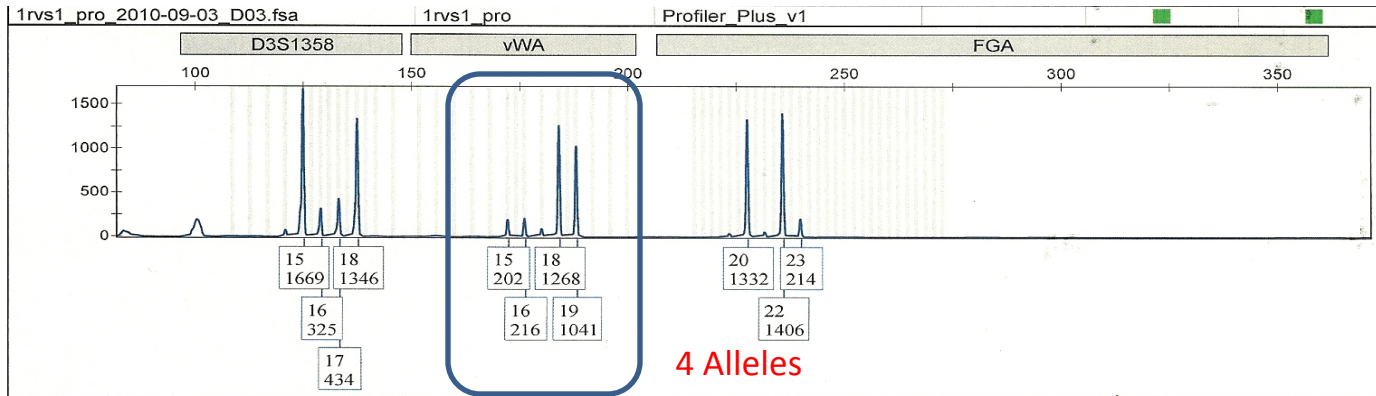
Formation of possible genotypes depends on PHRs allowed and the mixture ratio

$$\frac{PH_C + PH_D}{PH_A + PH_B}$$

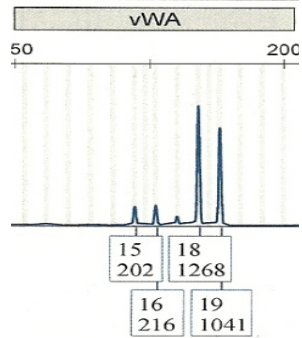
Sum major peak heights

Sum minor peak heights

Mixture Ratios



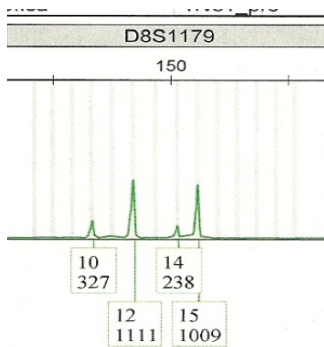
Calculate the Major/Minor Ratio



18 = 1268 rfu
19 = 1041 rfu PHR = 82%

$$(1268 + 1041) / (202 + 216) = 5.52$$

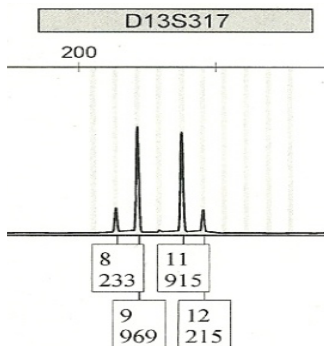
15 = 202 rfu
16 = 216 rfu PHR = 93%



12 = 1111 rfu
15 = 1009 rfu PHR = 90%

$$(1111 + 1009) / (327 + 238) = 3.75$$

10 = 327 rfu
14 = 238 rfu PHR = 72%

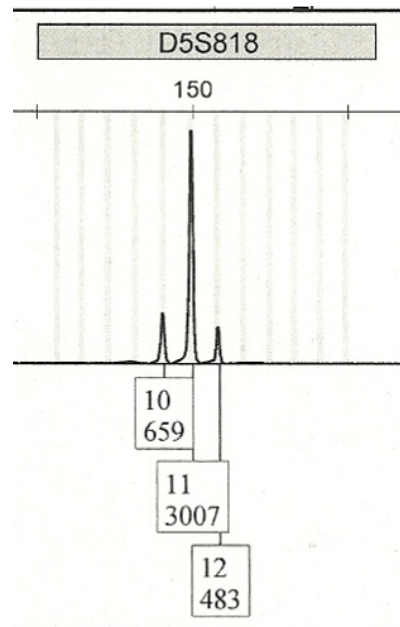


9 = 969 rfu
11 = 915 rfu PHR = 94%

$$(969 + 915) / (233 + 215) = 4.2$$

8 = 233 rfu
12 = 215 rfu PHR = 92%

Deconvolute a Locus with 3 Alleles



Possible Genotypes

- 10,10 and 11,12
- 11,11 and 10,12
- 10,11 and 12,12
- 10,11 and 11,12
- 10,11 and 10,12
- 10,12 and 11,12

Assuming 2 contributors, can any of these genotypes be ruled out?

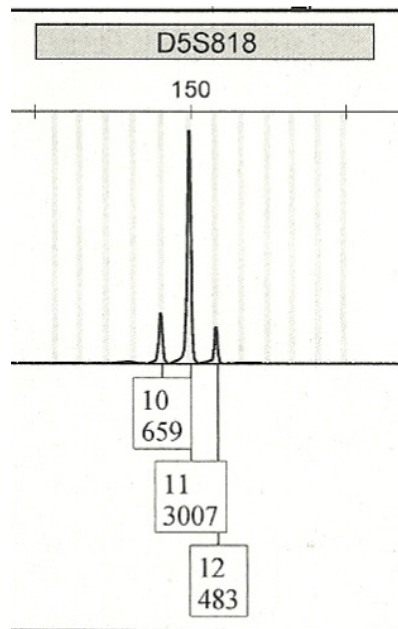
Use PHRs to examine each combination

But how to access for a possible shared 11 allele?

Calculate a Contributor Ratio

Contributor Ratio

Calculating contributor ratio may aid in deconvoluting a three allele locus



$$\frac{PH_{11}}{PH_{10} + PH_{11} + PH_{12}}$$

Sum major peak height(s)

Sum all peak heights

$$\frac{3007}{4149} = 0.72$$

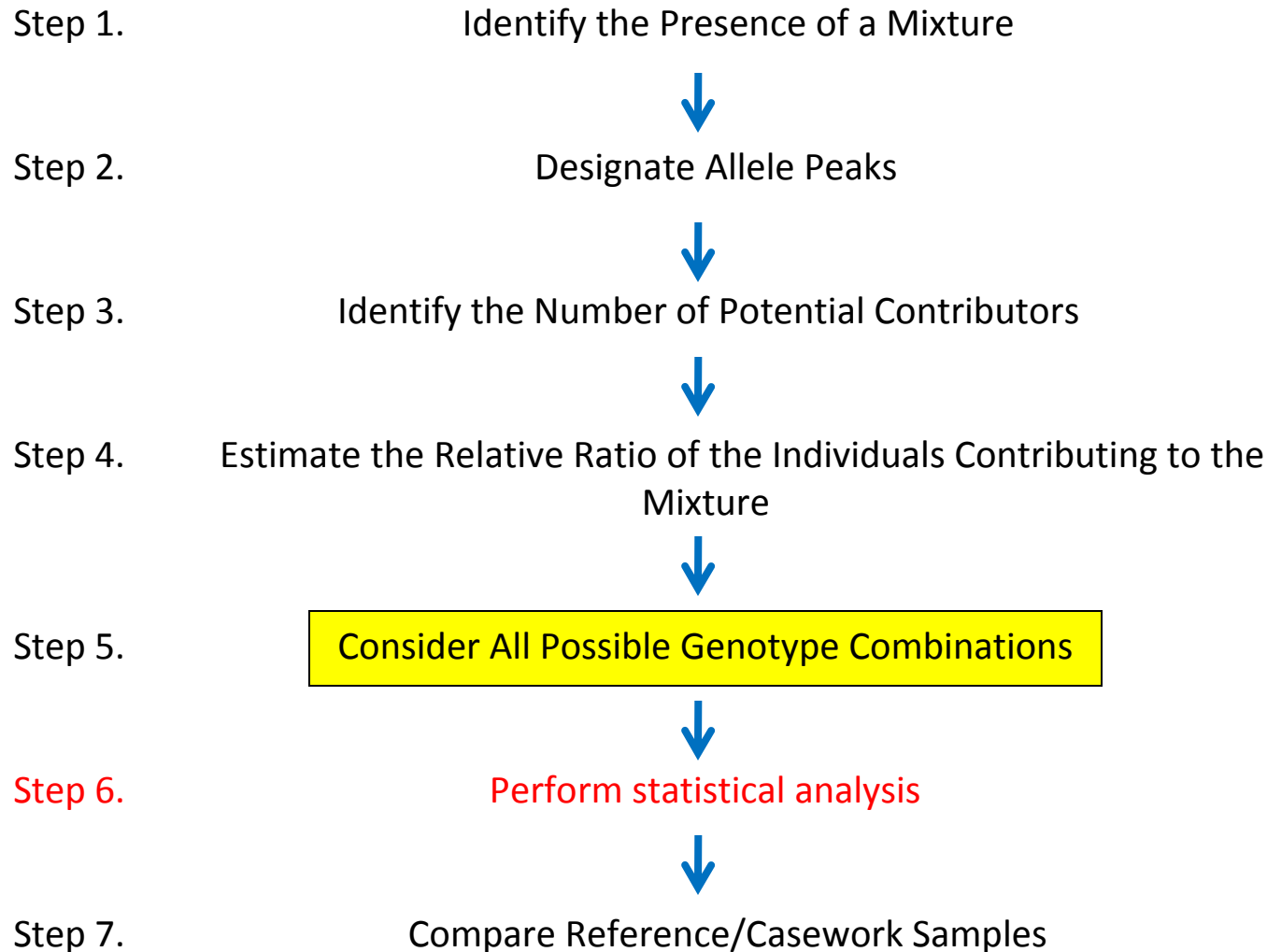
The contributor ratio of the 11 allele would be expected to be closer to 0.5 if it were the result of a mixture of a 10,11 and a 11,12

PHR of 10,12 = 0.73

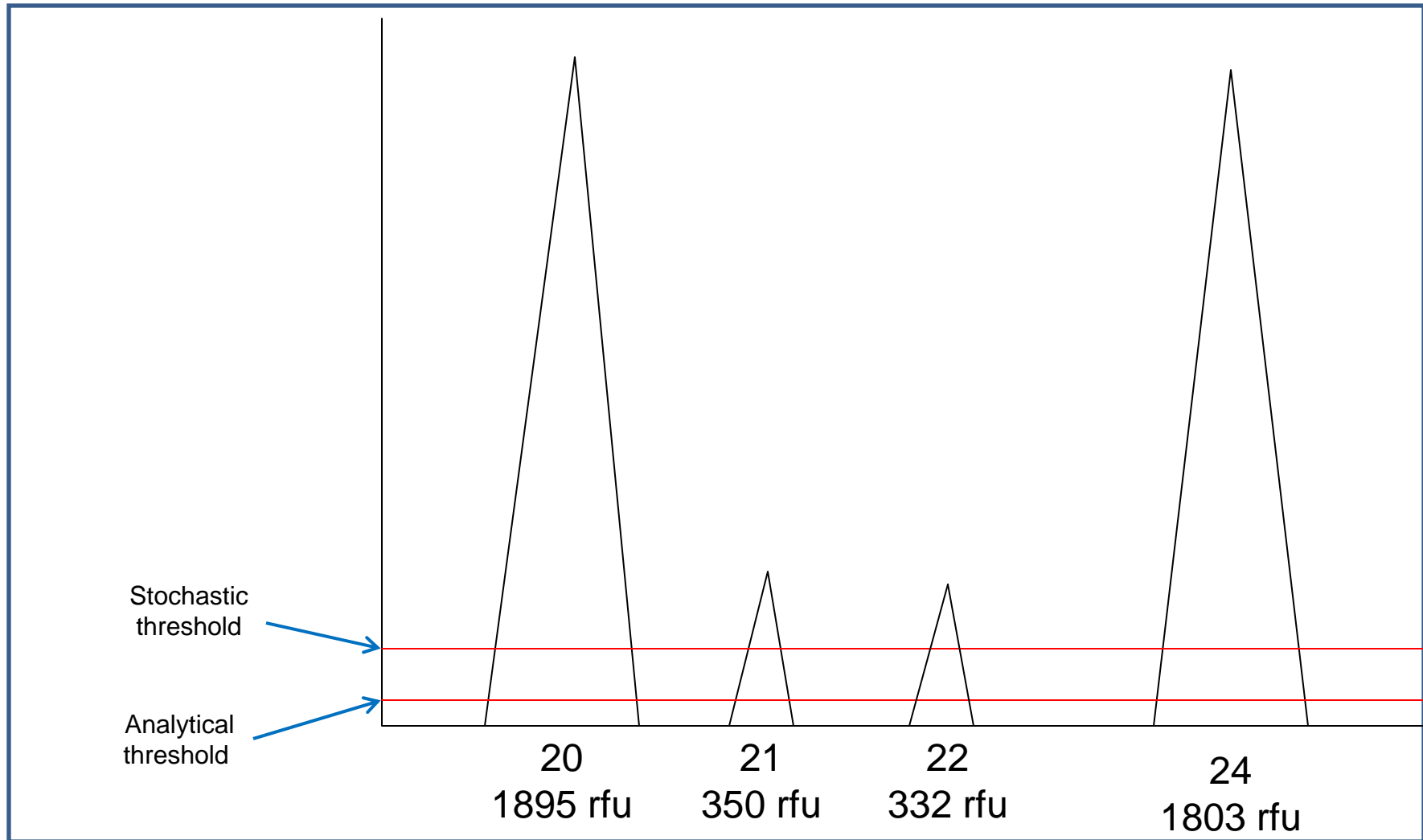
Supports a mixture of 10,12 and 11,11

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]

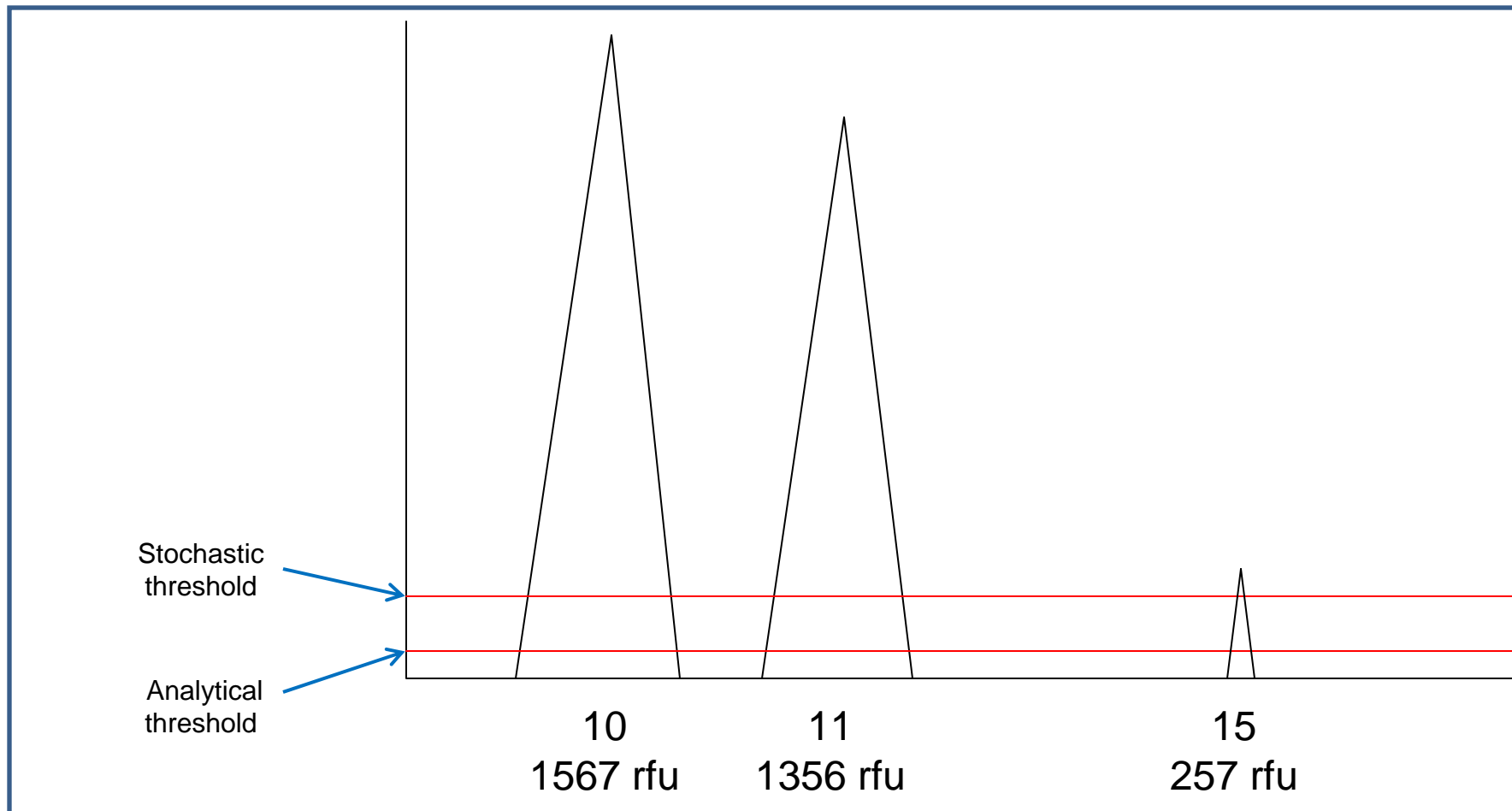


Resolvable for 2 contributors



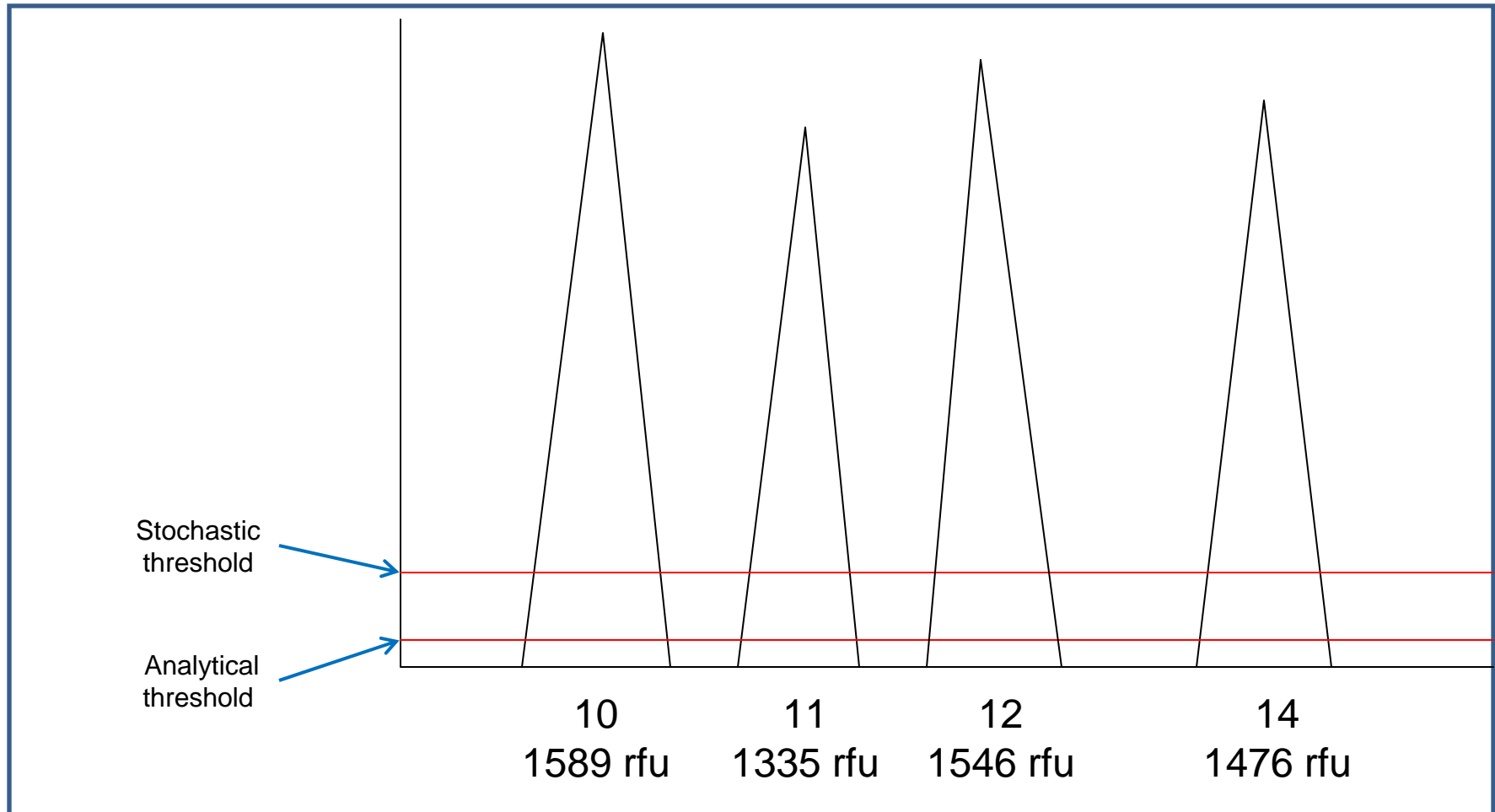
The mixture ratio is more than 4 if the genotypes are 20, 24 and 21, 22

Resolvable for major contributor only

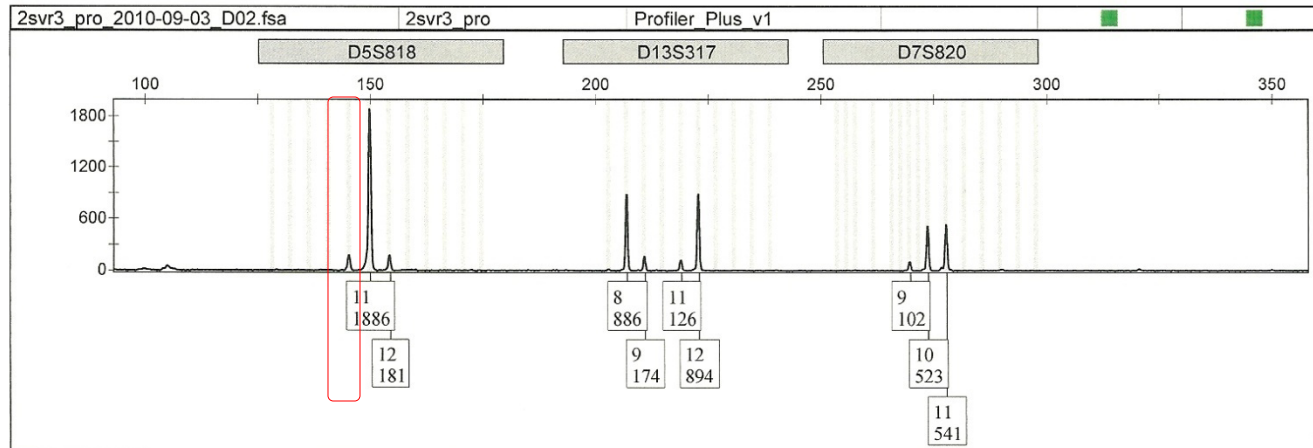
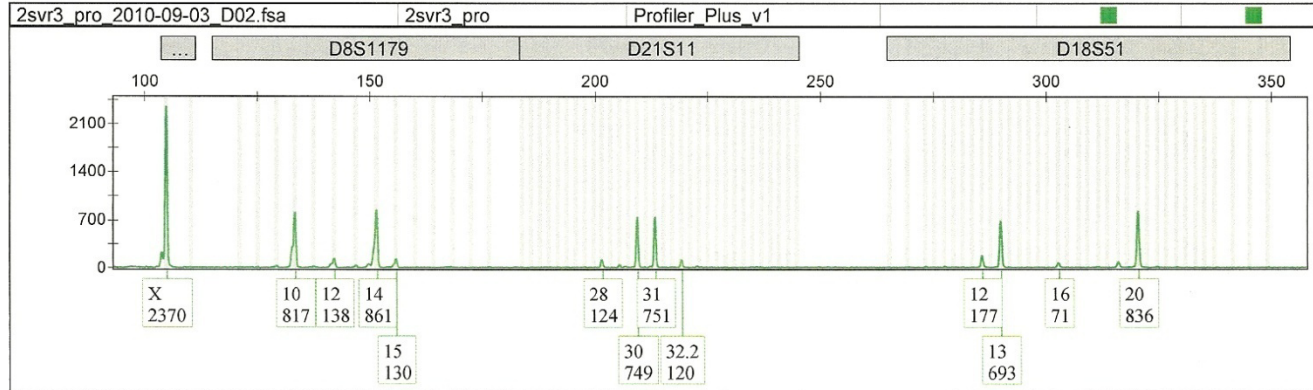
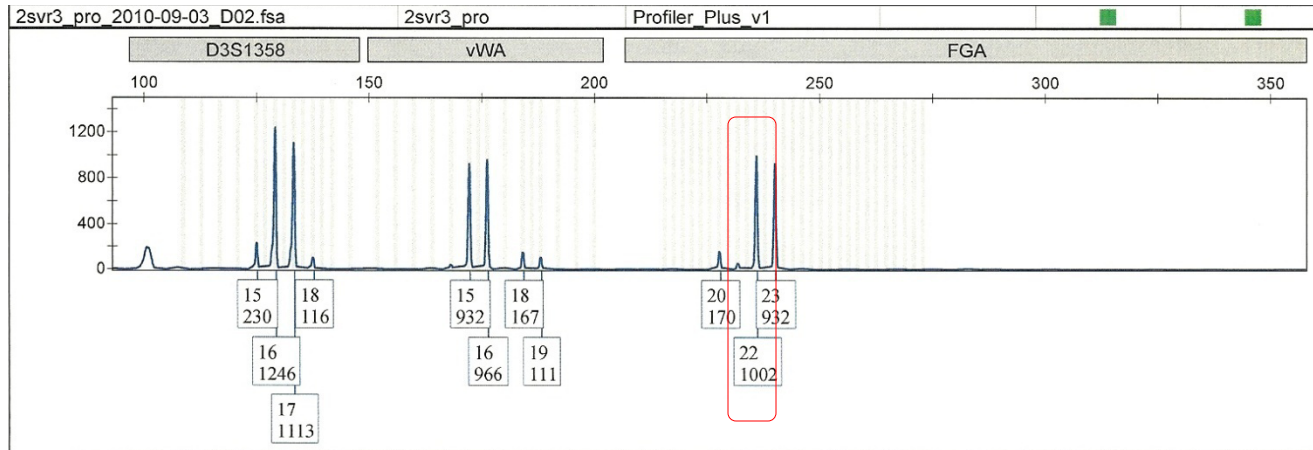


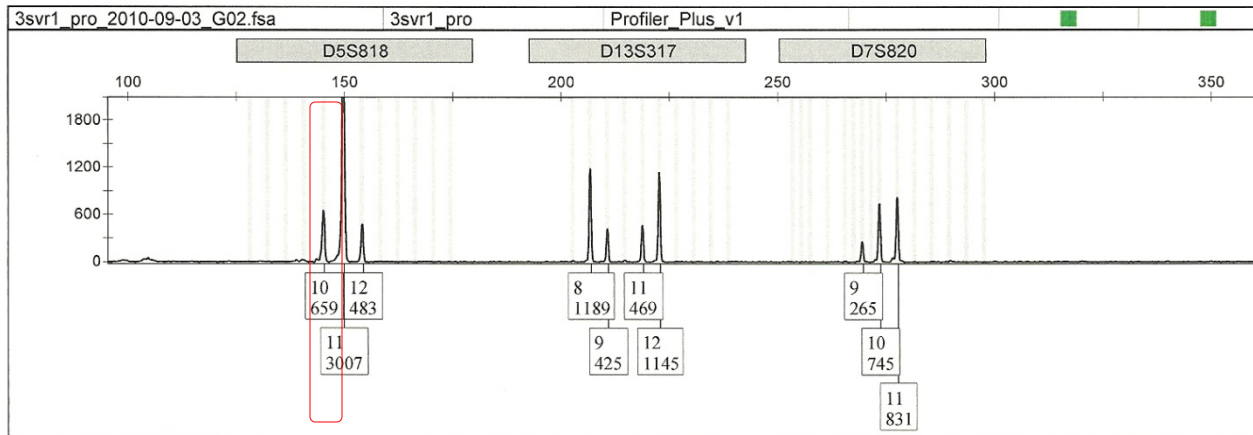
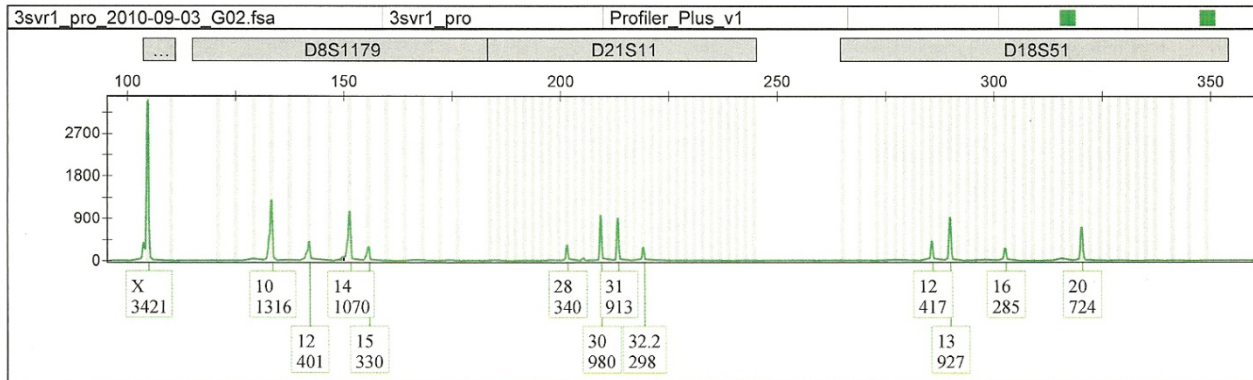
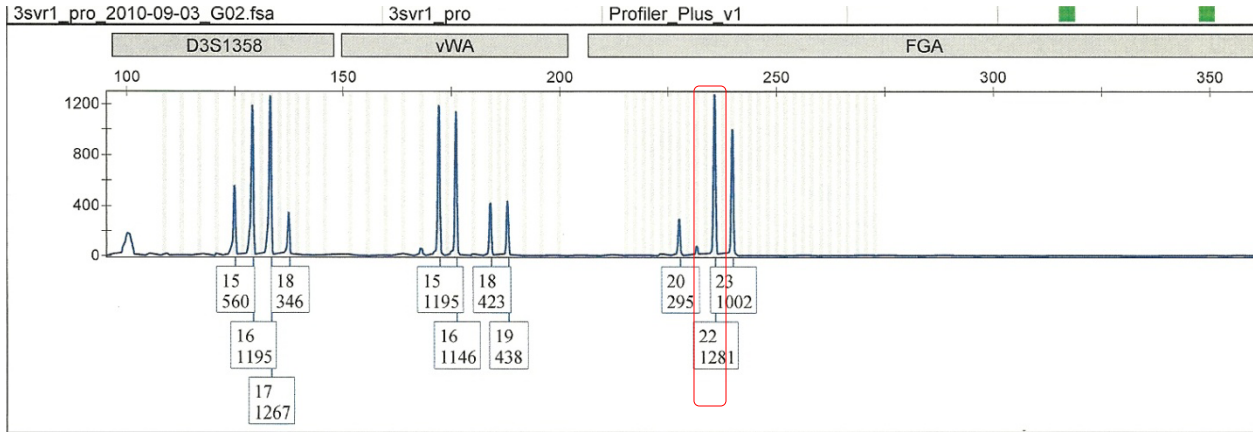
It is not clear if there may also be a minor 10 or 11 allele present, therefore only a major profile (10, 11) is unequivocally resolvable

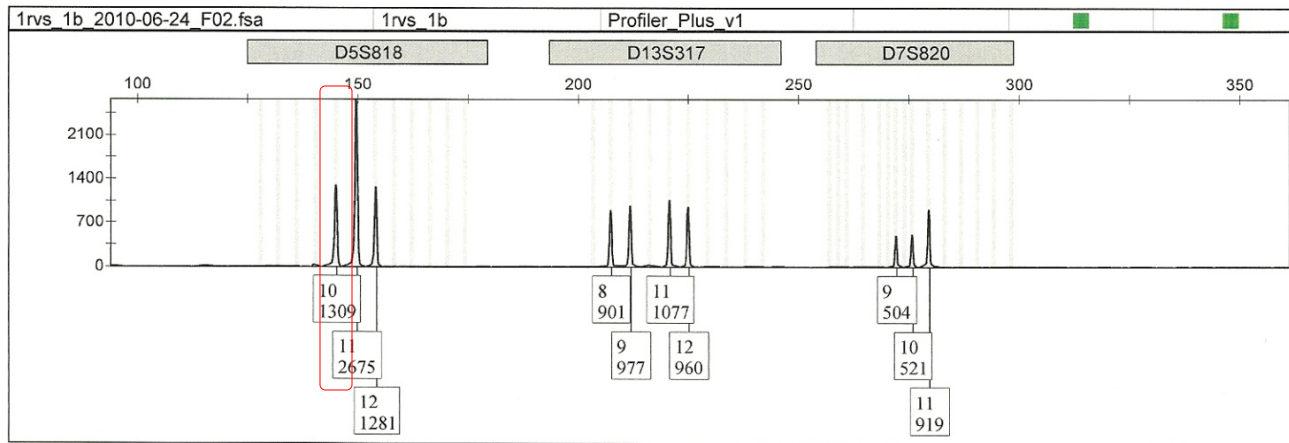
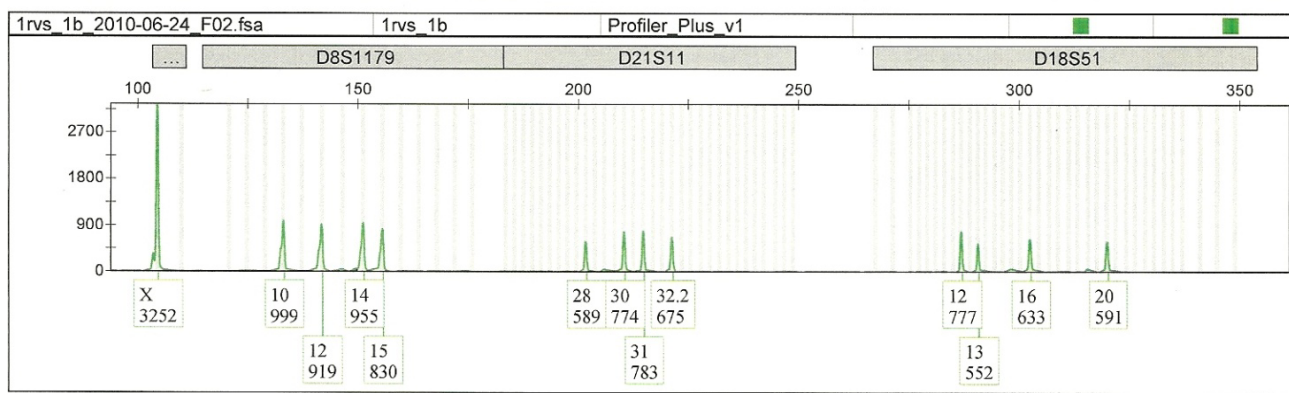
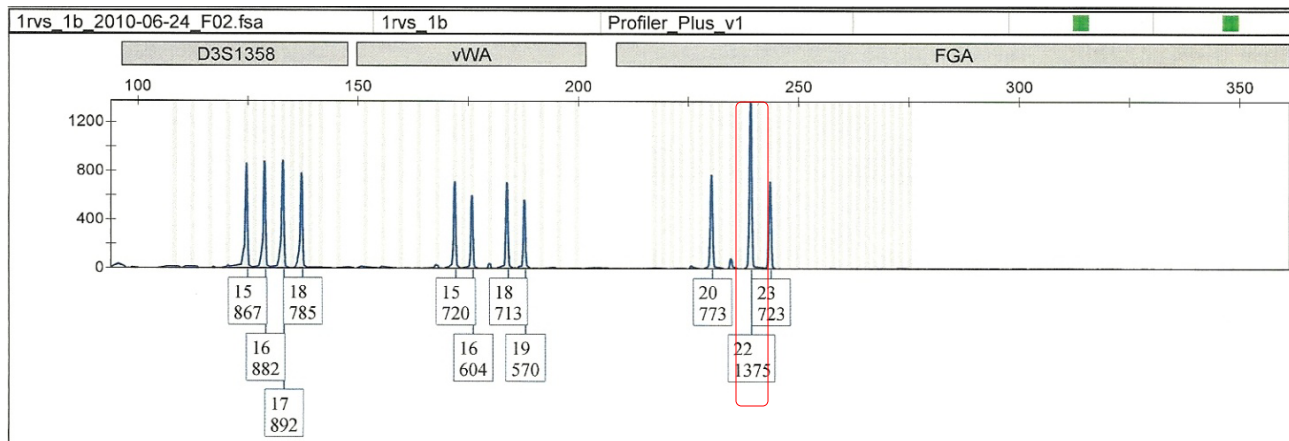
Not resolvable (no identifiable major/minor contributors)



None of the contributor peaks meet the 4:1 mixture ratio







Challenge to using PHRs for low level contributors

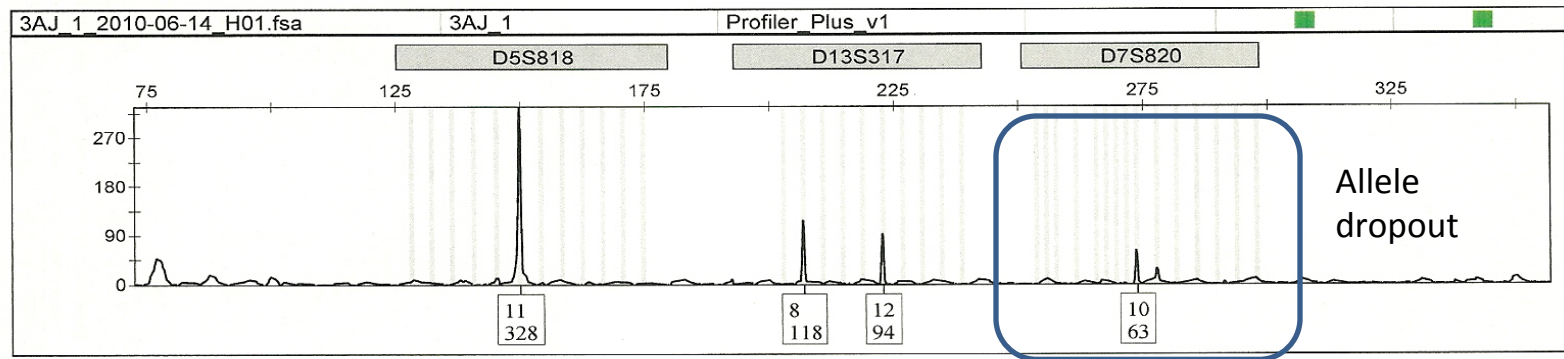
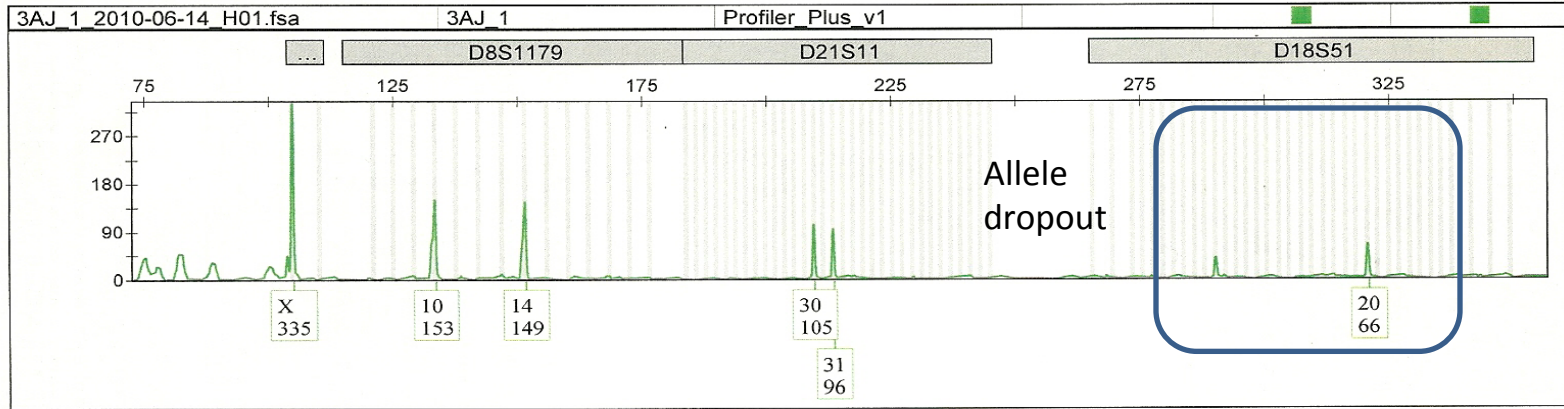
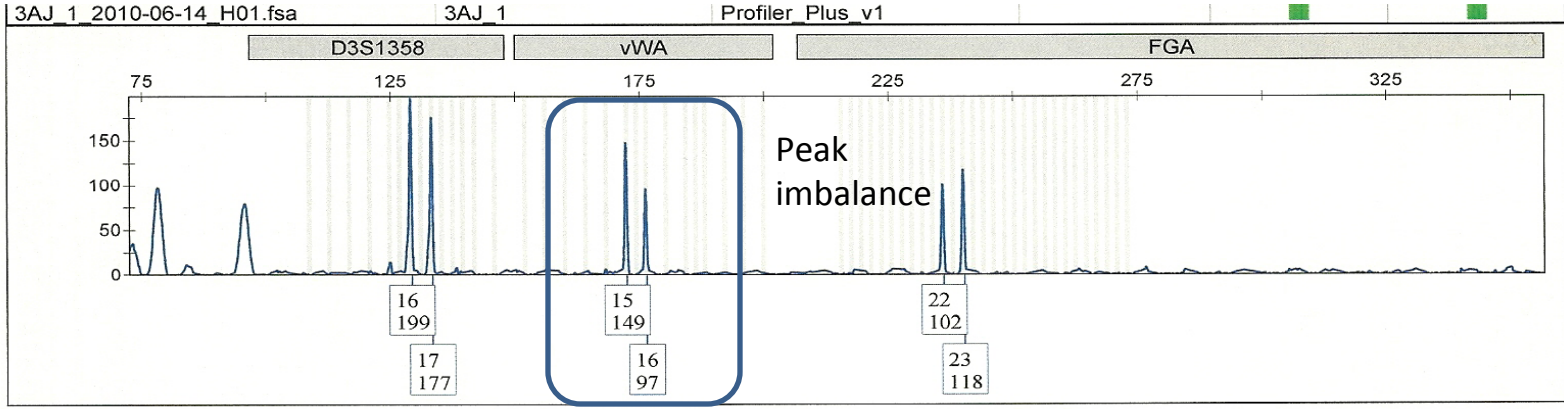
Even though the major contributors to a mixture may have optimal quantities of amplifiable DNA, the minor contributors can be in the stochastic range

PHRs can change significantly for contributors in the **stochastic range**

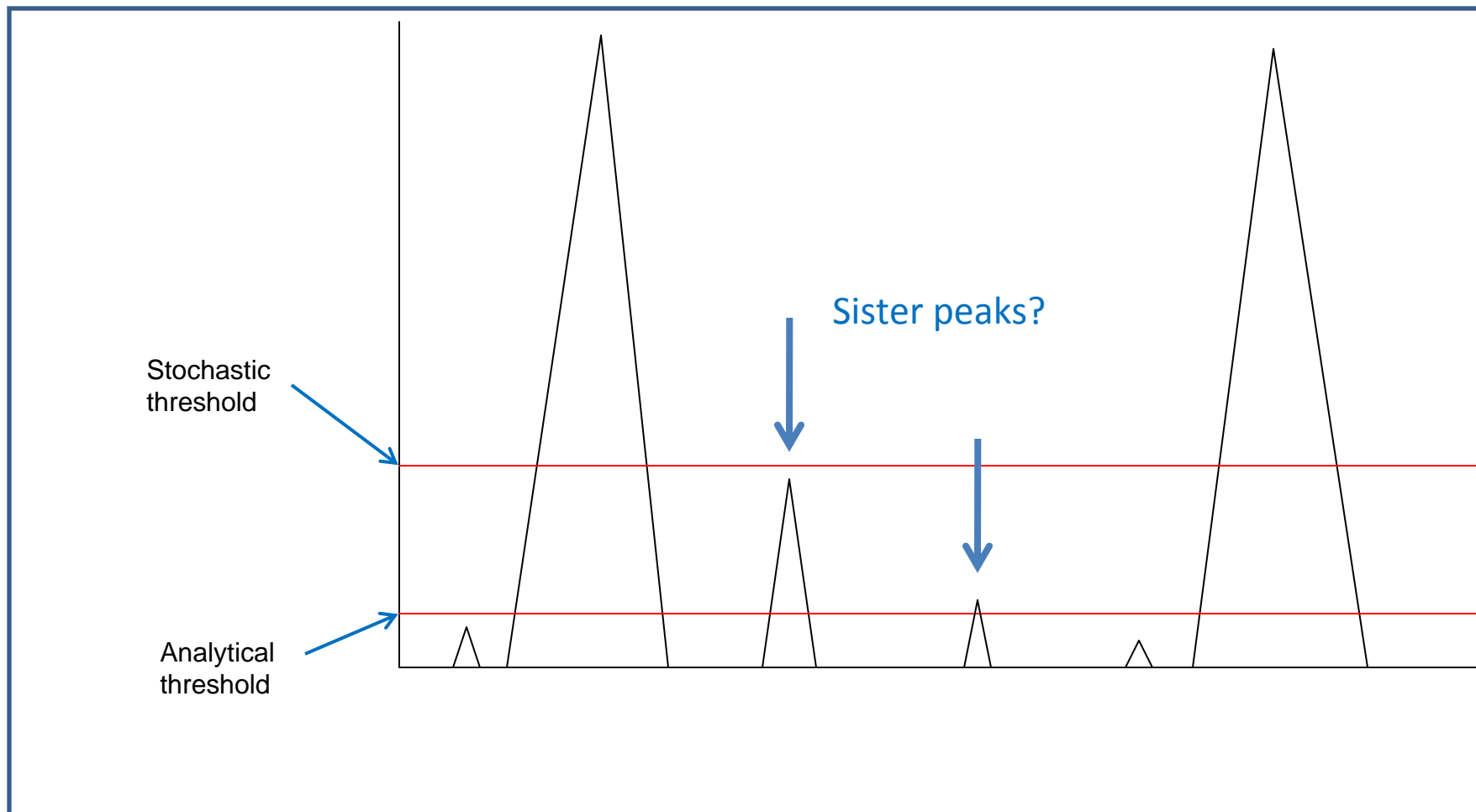
3.5.3.1. Differential degradation of the contributors to a mixture may impact the mixture ratio across the entire profile

Further validation may be useful to aid in interpreting samples that fall into this category

Relevant data can be gathered from sensitivity studies as well



Stochastic range Minor contributor



Minor Contributors

- It is possible (even likely) that a minor contributor's alleles may be masked by those of the major contributor(s)
- For this reason such alleles may not be unequivocally detectable or disregarded
- Determination of the minor contributor profile as a single source may be possible only at those loci where alleles are unequivocal or PHR data support only one possible profile for that minor contributor
- Number of contributors? Requires **assumptions**

Consider All Possible Genotype Combinations

Special consideration may be given to samples with a known contributor → **Intimate Samples**

SOP should clearly define what is an Intimate Sample

- swabs taken from a person's body (vaginal, oral, etc.)
- fingernail scrapings/clippings
- removed clothing (?)
- other personal items (?)

Under **defined** circumstances the profile of the known contributor may be subtracted out of the profile

Guideline **3.5.7** Mixtures with a Known Contributor(s)

Deducing a Second Contributor

Assuming two contributors in the mixture and one of them is known:

4 alleles observed:

- Known is heterozygous deduce 2nd person

3 alleles observed:

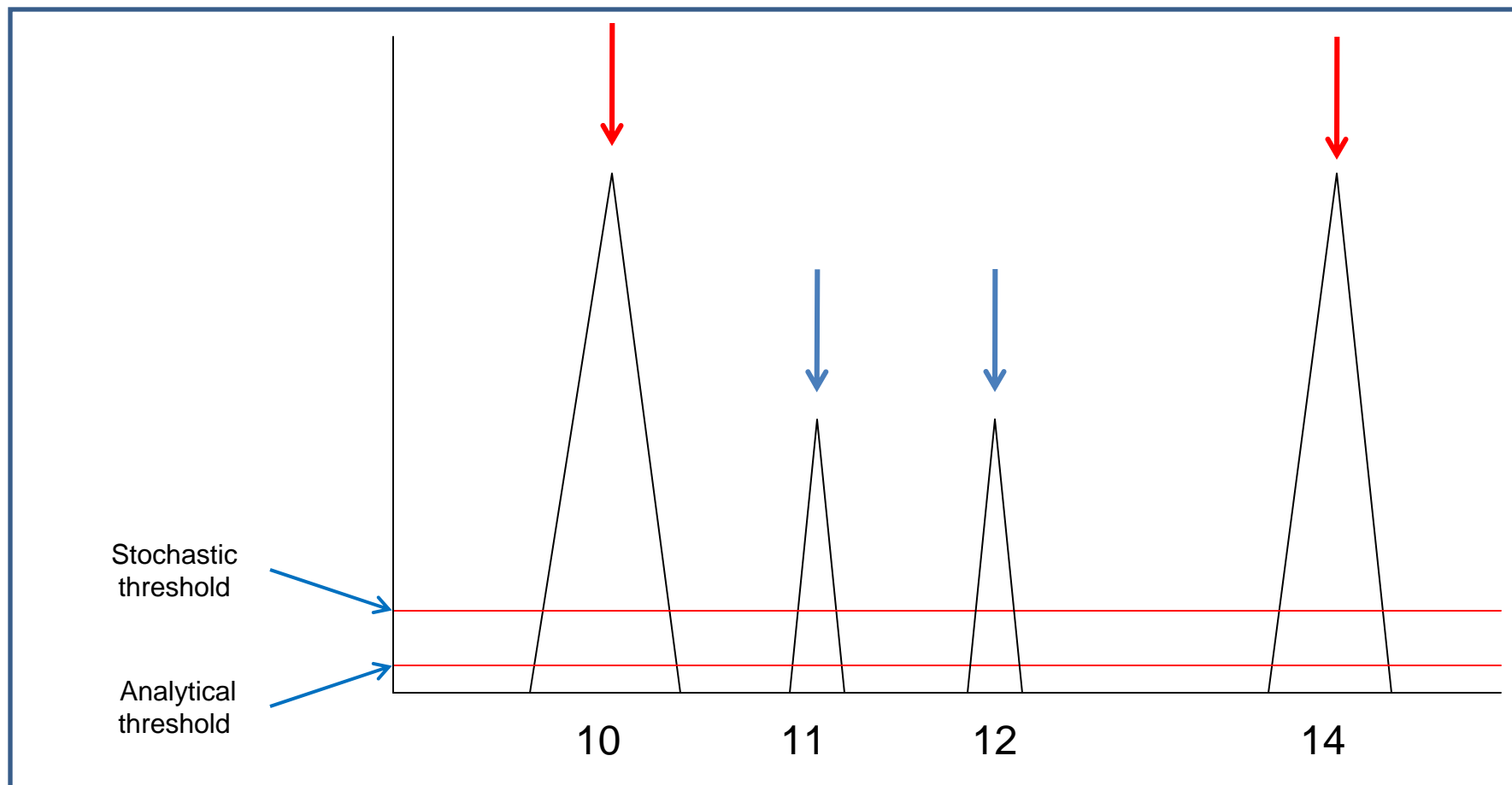
- Known is homozygous deduce 2nd person
- Known is heterozygous identify obligate allele of 2nd person

2 alleles observed:

- Known is homozygous identify obligate allele of 2nd person
- Known is heterozygous cannot deduce

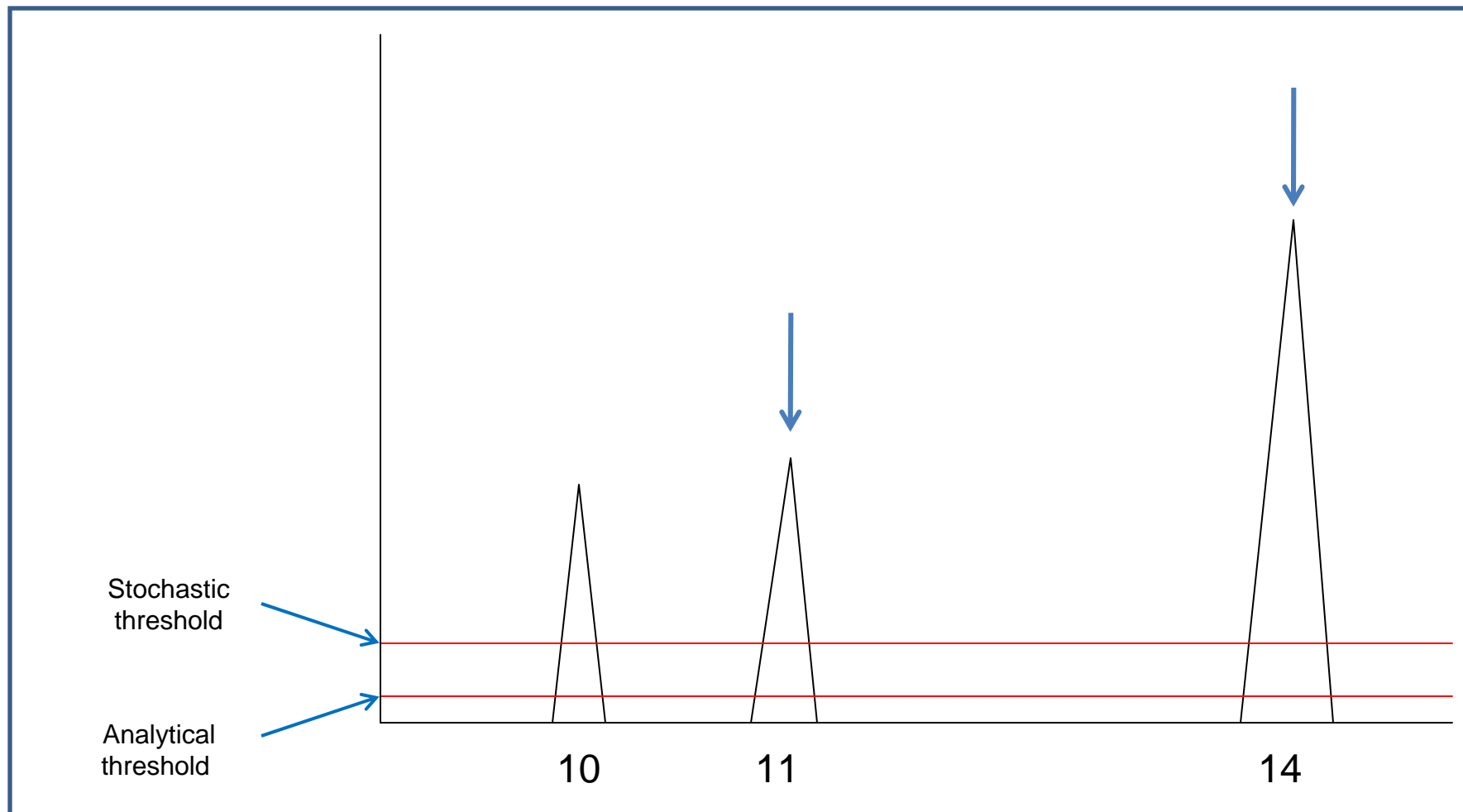
Consider All Possible Genotype Combinations

3.5.7.1. At a minimum, where there is no indication of sharing of the known and obligate alleles, the laboratory should separate out those alleles attributable to the known sample (e.g., victim, consensual partner, etc.).



Consider All Possible Genotype Combinations

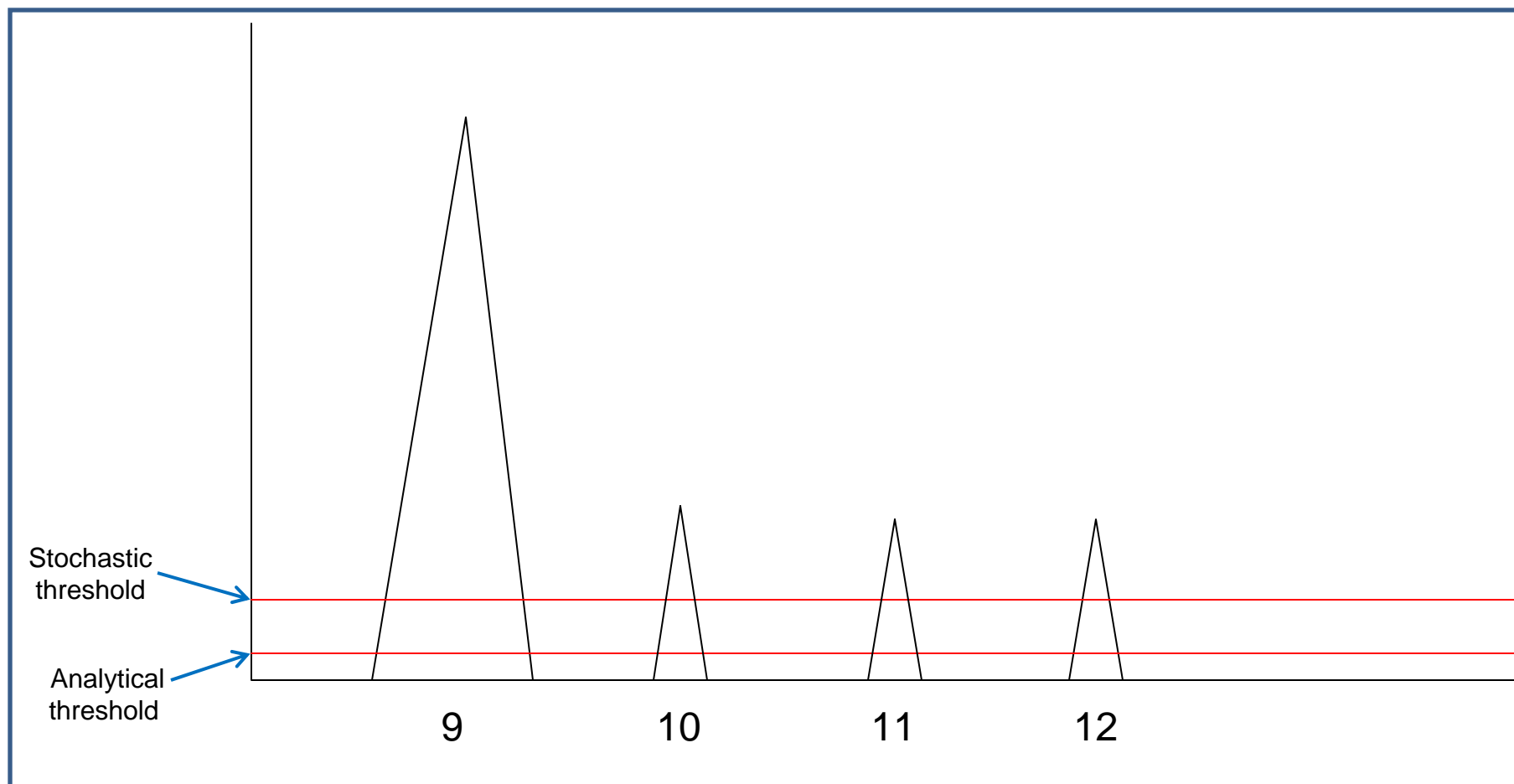
3.5.7.2. To further refine the obligate alleles in a profile, the laboratory may establish guidelines for addressing potential sharing of alleles among the individual known to have contributed to a sample and the additional contributor(s).



Additive Effects of Allele Sharing

- The larger the number of contributors to a mixture, the more allelic overlap is expected due to the sharing of alleles among contributors
- An allelic peak in a mixture could be the result of the combination of multiple copies of that allele from different donors
- The ability to assign specific genotypes directly or indirectly via deconvolution or subtraction will be lost

Additive Effects of Allele Sharing



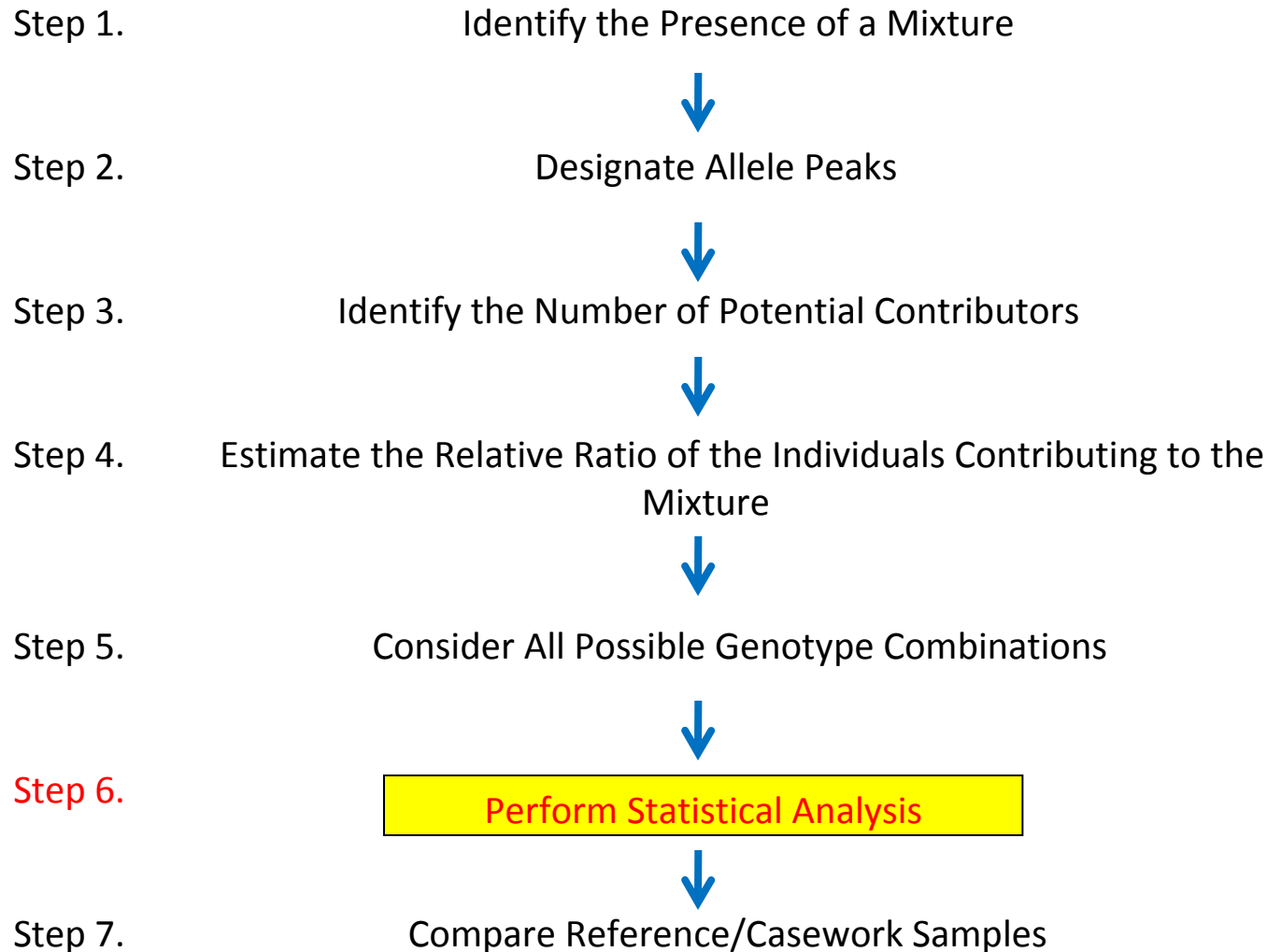
Minor contributors? Major contributors? Other loci may provide some information

Difficult Mixtures

- All of them?
- Some mixtures are not easily interpretable (or interpretable at all)
- Mixtures of three or more people which cannot be deconvoluted or subtracted
- Partial profiles
- Inclusions and exclusions can still be made, but they may not be appropriate for statistical analysis (stochastic range peaks)
 - 3.6.2.1. For partial profiles, the determination of which alleles/loci are suitable for comparison and statistical analysis should be made prior to comparison to the known profiles.
 - 3.6.2.2. The laboratory should establish guidelines for inclusions and exclusions when a known individual's DNA profile is not fully observed in the evidentiary profile.

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Perform Statistical Analysis

Guideline **3.6.1**. The laboratory must establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.

The statistical analysis should be done prior to comparison with the known samples (with the exception of subtracting out assumed contributors such as in intimate samples)

Significantly reduces analyst bias, real or perceived

Statistics are calculated for a profile NOT a comparison

Perform Statistical Analysis

Guideline 4.1. The laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.

Multiple tools

- Combined Probability of Inclusion/Exclusion (CPI/CPE)
- Random Match Probability (restricted/unrestricted RMP)
- Likelihood Ratio

The genetic loci and assumptions used for statistical calculations must be documented, at a minimum, in the case notes.

Identifying the Number of Potential Contributors and Statistical Analysis

Assumptions about the number of contributors will have significant impact on statistical calculations

- No assumptions on the number of contributors;

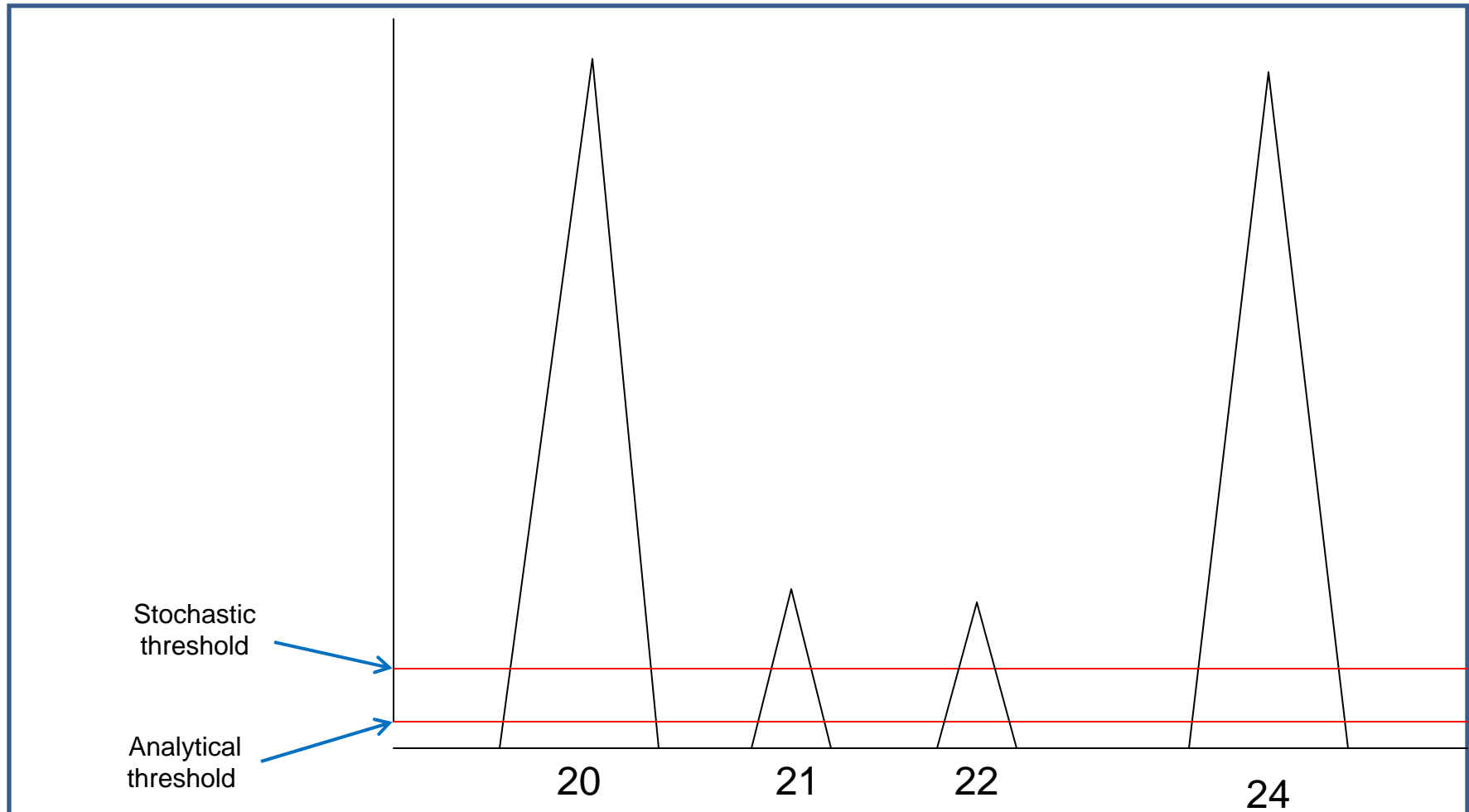
Unrestricted calculation \longrightarrow All combinations of alleles are deemed possible (relative peak height differences are not utilized)

- Assumption on the number of contributors;

Restricted \longrightarrow Based on the relative peak heights and mixture ratio assessments of an evidentiary profile, alleles are paired only where specific combinations of alleles are deemed possible

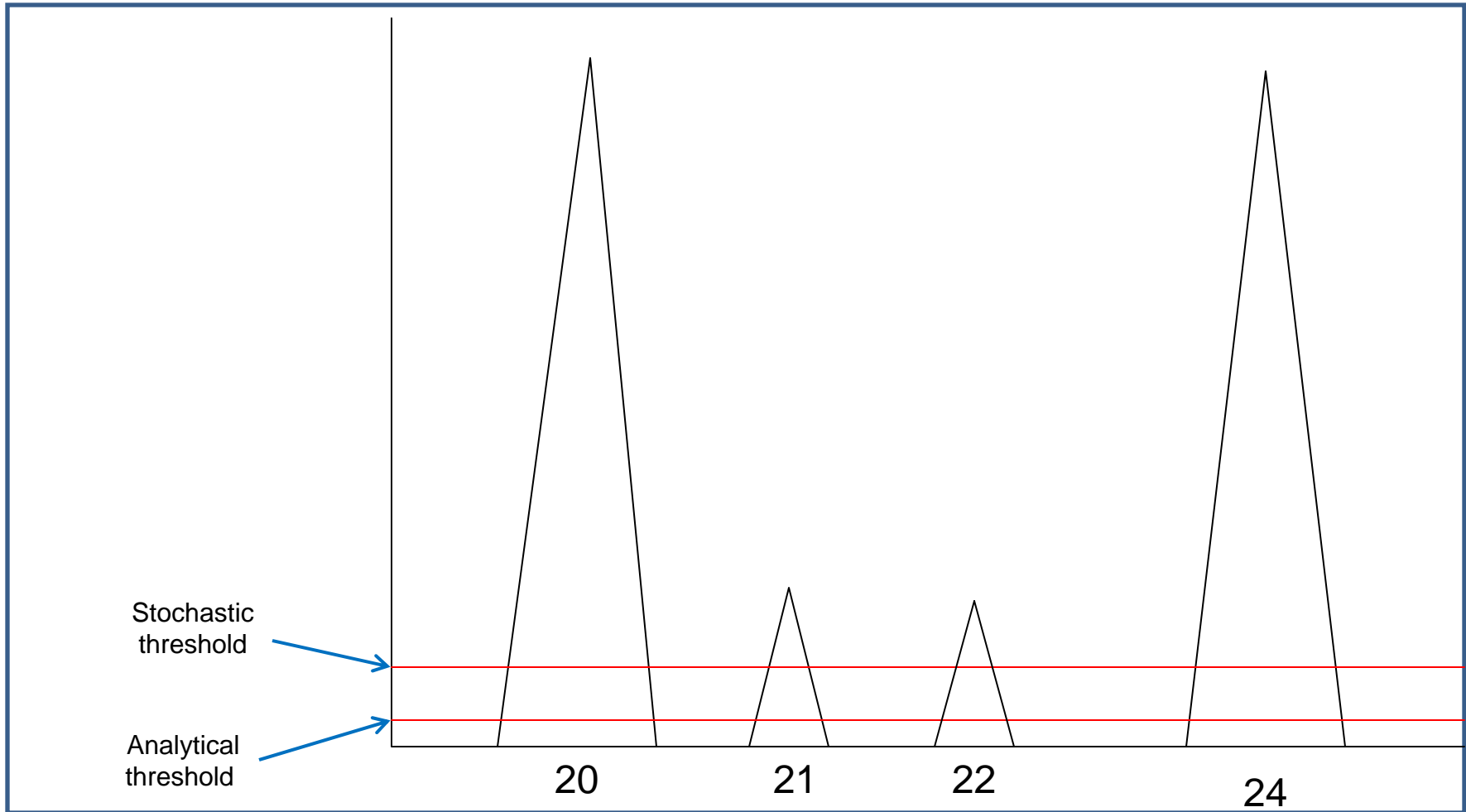
Guideline 4 Statistical Analysis of DNA Typing Results

Unrestricted



No deconvolution, all possible combinations considered equally; $20/21 + 20/22 + 20/24 + 21/22 + 21/24 + 22/24$

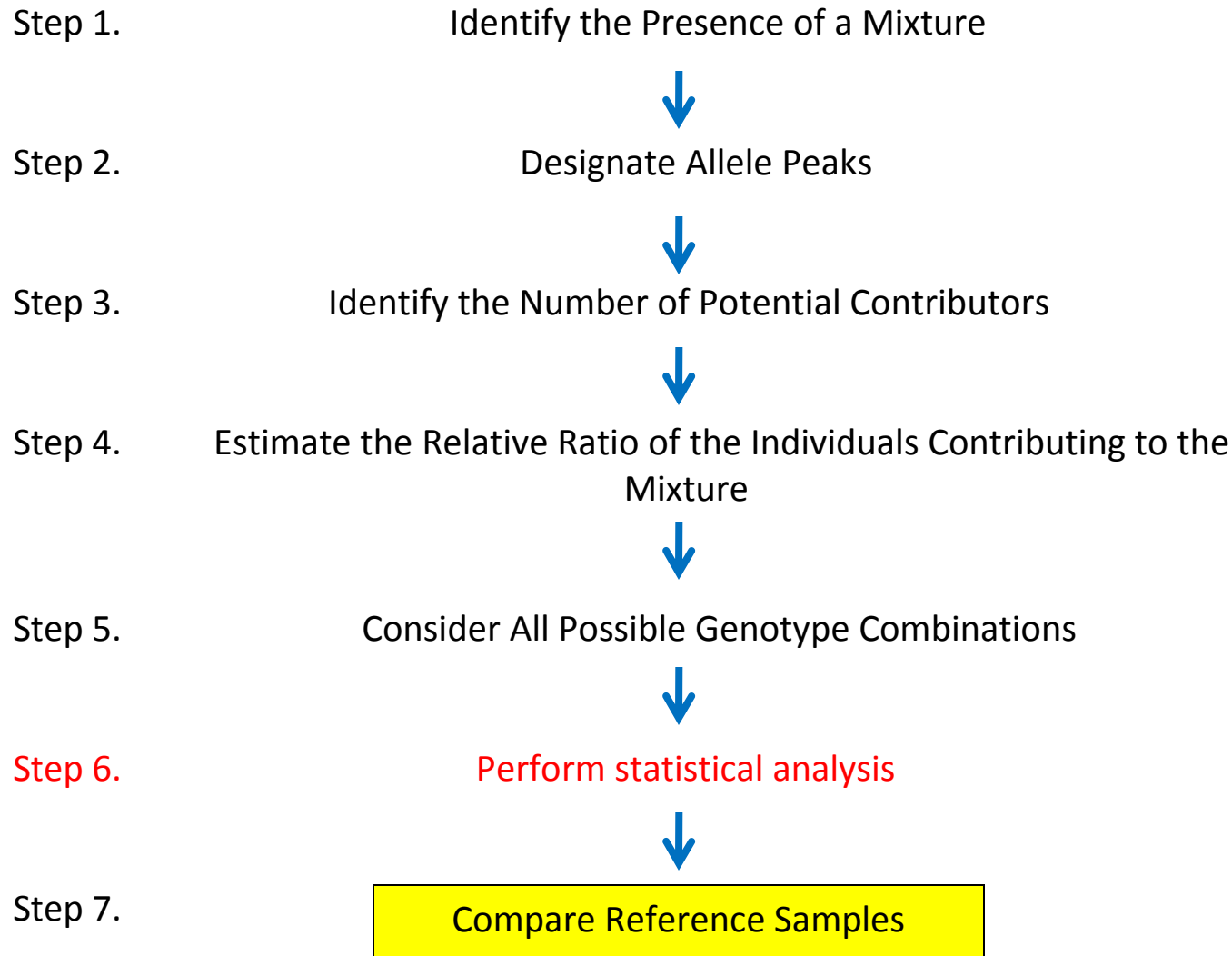
Restricted



Deconvoluted based on PHR and MR; ~~20/21~~ + ~~20/22~~ + 20/24 + 21/22 + ~~21/24~~ + ~~22/24~~

Steps in the Mixture Interpretation Process

[Adapted from Clayton *et al.* (1998) *Forensic Sci. Int.* 91:55-70]



Compare Reference/Casework Samples

Guideline 3.6 Comparison of DNA Typing Results

The following determinations can be made upon comparison of evidentiary and known DNA typing results (and between evidentiary samples):

- The known individual cannot be excluded (i.e., is included) as a possible contributor to the DNA obtained from an evidentiary item.
- The known individual is excluded as a possible contributor.
- The DNA typing results are inconclusive/uninterpretable.
- The DNA typing results from multiple evidentiary items are consistent or inconsistent with originating from a common source(s).

Compare Reference/Casework Samples

Guideline 3.6.4. For mixtures for which two or more individuals cannot be excluded as potential contributors, the laboratory may establish guidelines for assessing whether all of the DNA typing results obtained from the mixed sample are accounted for by the multiple known samples.

This guideline is useful for mixtures which cannot be deconvoluted or subtracted (Mixture Ratio < 3), however the reference samples can account for all of the observed alleles thus supporting an inclusionary determination

Compare Reference/Casework Samples

Guideline 3.6.6. The laboratory should establish guidelines for identifying DNA typing results for which comparisons of evidentiary and known samples are not made (at a minimum, to include inconclusive/uninterpretable results).

Inconclusive/uninterpretable

- Difficult to define
 - results at all loci, but at least 5 contributors at each locus?
 - results at only 1 locus?
 - indications of allele dropout/stochastic range?
- Often a source of analyst confusion
- The more clearly this category is defined, the better your **process** will be

Summary

Mixture interpretation is an involved process

- Use the guidelines as a start
- It is much less overwhelming when taken one step at a time
- Incorporate documentation into each step
- Use your validation data to the fullest extent
- There are multiple options for most of the steps
- Education and Training

Acknowledgements

- ❖ Abigail Jackson and Sarah Clark; UNH students who provided the mixture data
- ❖ Thank you for your attention
- ❖ Any questions/comments
madamowicz@newhaven.edu