1 RESEARCH ARTICLE

2 Calculating likelihoods and likelihood ratios at SNPs-based 3 mixtures. A reappraisal of the binomial inference, as applied to 4 forensic identity tests

5 Vincenzo L. Pascali

6 Department of healthcare surveillance and bioethics, Catholic University, Rome, Italy

7 Vincenzolorenzo.pascali@unicatt.it

8 Key words: mixtures; SNPs; genetic identification

9 Short title: **mixture analysis with SNPs**

10 Data Availability Statement: The MS Excel files cited in this article are freely available at:

11 www.vincepascali.it/docs/NITZsc_2_to_8_contrs.xlsx

12 www.vincepascali.it/docs/NITZq_SNP_2PM.xlsx

13 www.vincepascali.it/docs/NITZq_SNP_3PM.xlsx

14 www.vincepascali.it/docs/NITZq_SNP_4PM.xlsx

15

16 Funding: This work received no financial support

17 Competing interests: none

18

19 Abstract

20 Single nucleotide polymorphisms (SNPs) are useful forensic markers. When a SNPs-based
 21 forensic protocol targets a body fluid stain, it returns elementary evidence regardless of the
 22 number of individuals that might have contributed to the stain deposition. Therefore, drawing
 23 inference from a mixed stain with SNPs is different than drawing it while using multinomial
 24 polymorphisms. We here revisit this subject, with a view to contribute to a fresher insight into it.
 25 First, we manage to model conditional semi-continuous likelihoods in terms of matrices of
 26 genotype permutations vs number of contributors (NTZsc). Secondly, we redefine some algebraic
 27 formulas to approach the semi-continuous calculation. To address allelic dropouts, we introduce a
 28 peak height ratio index ('h', or: the minor read divided by the major read at any NGS-based
 29 typing result) into the semi-continuous formulas, for they to act as an acceptable proxy of the
 30 'split drop' (Haned et al, 2012) model of calculation. Secondly, we introduce a new, empirical
 31 method to deduct the *expected* quantitative ratio at which the contributors of a mixture have
 32 originally mixed and the *observed* ratio generated by each genotype combination at each locus.
 33 Compliance between observed and expected quantity ratios is measured in terms of $(1-\chi^2)$
 34 values at each state of a locus deconvolution. These probability values are multiplied, along with
 35 the h index, to the relevant population probabilities to weigh the overall plausibility of each
 36 combination according to the quantitative perspective. We compare calculation performances of
 37 our empirical procedure (NITZq) with those of the EUROFORMIX software ver. **3.0.3**. NITZq
 38 generates LR values a few orders of magnitude lower than EUROFORMIX when true contributors
 39 are used as POIs, but much lower LR values when false contributors are used as POIs. NITZ
 40 calculation routines may be useful, especially in combination with mass genomics typing
 41 protocols.

42 Introduction

43 In forensic genetics, a long-standing tradition of good practice has endorsed the adoption of multi-
 44 allelic polymorphisms (essentially, short tandem repeats; STRs) for the routine casework. The
 45 assumption underlying this choice is that a handful of these markers can assemble effective
 46 protocols and help to identify the donor(s) of a stain left at the crime scene. However, single
 47 nucleotide polymorphisms (SNPs) are recently qualifying as an prominent class of polymorphisms
 48 to adopt in the forensic work.

49 SNPs are mostly bi-allelic, and the Mendelian distribution of their genotypes (the algebraic
 50 expansion of a pair of gene frequencies) is easy to determine by just sampling a small population
 51 of unrelated individuals. SNPs are also ubiquitous, and a typical random human genome differs
 52 from the reference human genome by 4.1 - 5.0 million sites most of which are SNPs. [1]. *Last not*
 53 *least*, the next generation sequencing procedures (NGS) today used to type SNPs are not
 54 affected by most of those technical artefacts that normally complicate the process of STR typing.
 55 In view of this favourable background [2], several SNPs panels have been assembled for human
 56 identification and validated for the forensic routine [3-20].

57 However, the everyday application of these SNPs panels to cases of criminal identification has
 58 proceeded at slower pace than expected, possibly due to a series of circumstances:

- 59 a) SNPs protocols enlist more numerous markers compared to STRs; readapting calculation
 60 procedures / software routines (either based on the qualitative /semi-continuous approach
 61 [21] or on the quantitative approach [22-27]) to hundreds -thousands loci has proved to
 62 be more laborious and time consuming than initially expected;
- 63 b) Stains of forensic interest are mostly made of an admixture of multiple contributors; when
 64 a SNPs protocol is employed to type these mixed stains, the evidence tends to be uniformly
 65 biallelic and, based on this evidence, one cannot formally exclude any specific suspect. This

arises the serious problem of discriminating falsely compatible results from genuine matches by a statistic approach and not by the formal outcome of the test.

As a consequence of this deadlock, and with some notable exception [28] , the routine of criminal DNA tests still draws on without SNPs in several Countries.

The computational bottleneck at SNPs mixtures analysis has recently become the object of several studies. Estimates based on absolute likelihoods have been tried and proposed as the computational complement to mass typing protocols (Homer et al (2008) [29] ; Voskoboinik and Darvasi (2011) [30]), but this ‘frequentist’ approach (further developed by Isaacson et al (2015) [31]) has arisen criticisms (Clayton, 2010; [32]; Egeland et al 2012[33]). Furthermore, Voskoboinik et al 2015[34] have shown how a single contributor can be identified in mixtures by calculating true likelihood ratios at a protocol based on 3,000 SNPs ($0.002 < P(MAF) < 0.25$). Nearly at the same time, Gill et al (2015) [35] have explored the feasibility of the semi-continuous (or qualitative) calculations to the case of 2-person and 3-person mixtures. More recently, Bleka et al [36] have readapted the open-source *EUROFORMIX* software to the biallelic scheme and performed quantitative calculations starting from the ‘reads’ produced by a massive parallel sequencing (MPS) equipment.

In this paper we will revisit the logic of calculating binomial likelihoods and LR_s at forensic mixtures, under both the semi-continuous and the quantitative perspective.

Materials and methods

Semi-continuous calculations

Matrices to calculate likelihoods as at SNPs mixtures

At any mixture, the trace evidence (Et) contains n contributors ($n > 1$; this is the whole lot of individuals involved in a crime scene *and* leaving cells into the biological trace in matter). With

89 binomial SNPs, the evidence may only be single-allele ($E\{a\}$ or double-allele $E\{a,b\}$. When the
90 number of contributors increases, the double-allele evidence prevails and the mono-allelic
91 becomes an exception to find. If n is the only prior assumption to formulate on the mixture
92 content, $E\{a,b\}$ generates a matrix with as many columns as the number of contributors and as
93 many rows as the number of genotype permutations that can as equally express that evidence -
94 whereas $E\{a\}$ remains with no matrix at all. Such matrices have been denominated
95 ‘deconvolutions’. Deconvolutions of multinomial STR loci have been explored by Lynch and Cotton
96 (2018) [37] and these authors have warned on the numerosity and complexity of STRs
97 deconvolutions generated by dozens of genotypes and more than three contributors. This
98 warning, however, does not apply to SNPs: as opposed to STRs, SNPs have just one $E\{a,b\}$
99 deconvolution per contributor, no matter how numerous the contributors are. Moreover, the
100 number of records at each of these binomial deconvolutions remains acceptably low even the
101 number of contributors is high. In fact one can count 7 permutation rows (or deconvolution states)
102 at the 2PM case, 25 rows at the 3PM deconvolution, 79 at the 4PM, 249 at the 5PM, 727 at 6 PM,
103 2185 at 7PM and 6559 at 8PM. The two circumstances (one matrix per class of mixture; a limited
104 number of coexistent combinatorial records per matrix) make it convenient to pre-compile a
105 handful of universal matrices (one per number of contributors), place them into as many Excel
106 worksheets and use the worksheets for computing over all binomial loci of interest. Matrices may
107 be made universal by adopting always the same allele nomenclature - A for the quantitatively
108 predominant allele, B for the other; AA, AB and BB for the relevant genotypes - for whatever locus.
109 If used for semi-continuous calculations, matrices will only contain genotype frequencies, so that
110 using the universal matrix for the n^{th} SNP is only a matter of updating the worksheet’s pair of
111 gene frequencies with those of the locus of interest. Universal matrices become convenient to use

112 when the number of SNPs placed at a forensic protocol is counted in hundreds [17-20] or possibly
 113 thousands [16, 30, 34]).

114 When further prior hypotheses are formulated introducing η persons of interest (POIs; known
 115 contributors, or K_s ; $1 \leq \eta \leq n$) the inference to draw becomes conditional to both n and η
 116 (Benshop et al 2015; [38]). Just as noticed for E_t , sets of POIs can have a single-allele pattern $K_t\{a\}$
 117 or a biallelic pattern $K_t\{a,b\}$. POIs sets are used as a condition to single out subsets of the general
 118 matrix containing their genotype (see Appendix). This may be achieved by embedding a few,
 119 simple 'if-then' excel conditional formulas into the relevant Excel worksheet. These subsets are
 120 themselves true likelihoods and their arithmetical value is inferred by summing the permutation
 121 states products contained at only those records that contain the known genotype. We here supply
 122 seven precompiled spreadsheets (to be downloaded from https://www.vincepascali.it/docs/NITZsc_2
 123 [to 8 contrs.xlsx](#)), one per number of contributors (from 2PM to 8PM), each allowing computations
 124 that allocate up to two known genotype per matrix.

125 Calculating likelihoods by algebra

126 When compared one to another, the evidence and the POI allelic subsets generate nine
 127 assortments, shown in Fig 1 . Fig 1 contains matching cases and exclusions. Exclusions only
 128 emerge when the evidence is made of one-allele. When this happens, either K and E_t are
 129 disjointed sets ($K_t \cap E_t = \emptyset$; no allele in common; for example $E_t\{a\}$ comparing to $K_t\{b\}$) and
 130 viceversa) or E_t is a subset of K_t : $E_t \subset K_t$ (whereas in the case of compatibility one would
 131 normally expect the reverse $K_t \subset E_t$ to be true).

132 All of these nine cases may be worked out by algebra. Principles ruling the algebra are:

- 133 1. at a single-source trace, the whole evidence is: $P(E_t) (a+b)^2 = P(E_t\{a\}) + P(E_t\{b\}) + P(E_t\{a,b\}) = 1$;

2. the generalization of the single-source trace model to n contributors requires exponentiation to n of all members of the evidence: $P(a+b)^{2n} = P(a^{2n}) + P(b^{2n}) + P(a-b)^{2n}$ [21];
3. introducing η obliged contributors and calculating relevant likelihoods involves to subtract η to the original exponent: $P(a+b)^{2n} \rightarrow P(a+b)^{2(n-\eta)}$.

Fig 1. Assortments of the evidence and the POI allelic subsets.

This matrix shows that there is a limited number of outcomes (just nine, in fact) to expect at a SNPs based identity test.

Algebraic notations for compatible semi-continuous likelihoods

The following conditional formulas apply to compatible cases of table 1:

$$P(Et\{a\} | K\{a\}, n, \eta) = a^{2(n-\eta)} \quad (1)$$

$$P(Et\{b\} | K\{b\}, n, \eta) = b^{2(n-\eta)} \quad (2)$$

$$P(Et\{a,b\} | K\{a\}, n, \eta) = 1 - a^{2(n-\eta)} \quad (3)$$

$$P(Et\{a,b\} | K\{b\}, n, \eta) = 1 - b^{2(n-\eta)} \quad (4)$$

$$P(Et | K\{a,b\}, n) = 1 - a^{2n} - b^{2n} \quad (5)$$

$$P(Et | Et\{a,b\}, K\{a,b\}, n, \eta) = (a^2 + b^2 + 2ab)^{(n-\eta)} = 1 \quad (6)$$

where a, b in the right side of the equations are the gene frequencies of the homonymous SNPs alleles. These formulas are a generalization of those reported by Voskoboinik et al 2015 [34] – modified by introduction of rule 3 of previous chapter. As previously noticed by these authors, the six formulas do not account for cases where there is a mismatch between the evidence and the POI.

Algebraic notations for incompatible likelihoods

We here assume that the mono-allelic evidence derives from an originally bi-allelic evidence by the loss of one allele (allele dropout). Within our nomenclature system, only the minor quantity

allele can drop out of an evidence, with some probability ($P(D)$) that proportionally increases as soon as the template concentration decreases down to critical values. In the available literature $P(D)$ values have been inferred from experiments conducted on various templates dilution, by logistic regression analysis (Tvedebrink et al 2009 [39] ; and Buckleton et al 2014 [40]) – and these experimental values have been incorporated into semi-continuous calculations. A current computational approach relying on such dropout probabilities is the ‘split drop approach’, developed by Haned et al in 2012 [41] and readapted to the case of SNPs by Voskoboinik et al [[34]. The split-drop approach is actually adopted by the two widespread software LRMIX and EUROFORMIX [42].

We instead take the approach of incorporating dropout probabilities into a few algebraic formulas working out the task that is normally committed to the split drop approach:

$$P(\text{Et}\{a_d, b\} \mid K\{a\}; P(h); n; \eta) = h^2 * a^{2(n-\eta)} \quad (7)$$

$$P(\text{Et}\{a, b_d\} \mid K\{b\}; P(h); n; \eta) = h^2 * b^{2(n-\eta)} \quad (8)$$

$$P(\text{Et}\{a; b_d\} \mid K\{a; b\}; P(h); n; \eta) = h \quad \text{or:}$$

$$P(\text{Et}\{a_d, b\} \mid K\{a; b\}; P(h); n; \eta) = h \quad (9)$$

It is here to notice that, with reference to the notations adopted in figure 1, $\text{Et}\{a; b_d\}$ replaces the original $\text{Et}\{a\}$ notation and $\text{Et}\{a_d; b\}$ replaces $\text{Et}\{b\}$, respectively. $P(h)$ is a proxy for the experimental logit values. For its meaning, read on the following paragraphs.

Algebra for ‘dropped out and still compatible’ cases

Although the concept of dropout is invoked to account for cases of contributors that do not match the evidence, ‘exclusion’ and ‘dropout’ are not logically equivalents. In fact there may be instances where an allele drops out of the evidence and the resulting reference remains compatible to the POI. Such ‘latent’ dropout cases can be modelled by turning equations 1 and equ. 2 into:

$$P(Et\{a, b_d\} \mid K\{a\}, n, \eta) = (1-h) * a^{2(n-\eta)} \quad (1d)$$

$$P(Et\{a_d, b\} \mid K\{b\}, n, \eta) = (1-h) * b^{2(n-\eta)} \quad (2d)$$

It is to note that, if using our conventional nomenclature (a, major read allele; b, minor read allele) equation 2d is never actually used.

Complete algebraic inference

The algebraic solutions we have just introduced (1d; 2d; 3-6; 7-9) can now cover all the inferential patterns expected to show up at any test based on binomial markers. An overview of the nine relevant cases is in Fig 2.

Fig 2. Algebraic solutions given to all possible evidence/POI allelic subsets assortments of Fig1.

The mono-allelic evidence appearing in Fig 1 is reinterpreted under the assumption that dropout events may have altered $Et\{a, b\}$. Compatible cases are highlighted in yellow background, exclusions in green background.

The h index as a proxy for experimental P(D) values

Dropout occurs at SNPs too [43], but specific logit dropout values have not been so far exploited for SNPs. This is not surprising: these values should be inferred (by experimental data logit treatment) from *each locus in use at a given protocol- possibly by each laboratory adopting the protocol*. And because forensic protocols are often made of hundreds / thousands loci, this goal is not easy to attain in practice. That being the case, a good deal of semi-continuous calculations at SNPs mixtures (as those performed by the typical users of LRMIX and EUROFOMIX) end up to rely on arbitrary P(D) values. This encourages to search for other solutions.

We found the a convenient proxy for the P(D) is the h index, introduced by Kelly et al [44]), in its easiest form (the minor read simply divided by the major read). There are several reasons to introduce this index in place of logit-derived P(D) values:

a) h applies as well to single-source stains and mixtures;

- b) at mixtures, h may represent the whole propensity of the minor-read allele to drop out of the mixed evidence;
- c) $(1-h)$ can estimate the probability of existence of the two-allele evidence;
- d) The equivalence $h \approx P(D)$ has a theoretical justification: it has been predicted that a $P(D)$ can automatically emerge from the peak height analysis, without any external logistic regression modelling (Cowell, 2018 [45]); if starting from this standpoint, a typical $P(D)$ value to deduct from the peak area analysis cannot be but the h index.

Index h enters into the likelihood formulas as shown in the previous paragraphs.

Dropins

According to an official definition issued by the ISFS commission [46], dropin is an allele that is not associated with the crime sample and remains unaccounted both under the prosecutor and the defence hypothesis. In principle, a dropin phenomenon is nothing more than an additional, 'uninteresting' contributor to a mixture and its case could be modelled by increasing the number of contributors. Since the SNPs-based evidence generates relatively simple deconvolutions even at high numbers of contributors we address the dropin case by systematically switching the current scheme of computation to one scheme with one more contributor. We will later see that this has limited effect on the LR value.

Number of contributors

Guessing on the number of contributors that enter in a SNPs mixture may be difficult. The importance of the whole issue is directly proportional to the impact the value of n may have over the LR values. Bleka et al 2017 [36] have explored this issue by simulating 100 mixtures with two to up to six virtual contributors and calculated likelihoods at various quantitative POI proportions to the POI. They found that, when the POI is not quantitatively predominant, their two-person mixture model of analysis would return them higher LRs compared to the three- or more-person

models. They also found that, from three-persons mixtures onwards, increasing the number of contributors has a negligible effects over the POI's LR, regardless of the POI's quantitative predominance. We have an explanation to this phenomenon: unlike higher-rank mixtures, 2-persons mixtures cannot have permutation states with full assortment of binomial genotypes (AA+AB+BB). Switching from 2PM to 3PM has therefore the effect of completing the genotype assortment, elongating the deconvolution, increasing the number of states and lowering the MR. The same does not occur while switching from 3PM to 4PM or to higher-rank models. There is consequently much logic in believing that 3PM is the minimum standard to endorse at binomial mixture analysis. In this paper, however, we will always calculate LRs under three distinct hypotheses (2PM, 3PM, 4PM).

LRs

Each likelihood in figure 2 may represent the viewpoint of the prosecution or the defense at a typical judicial test. A ratio between two of these likelihoods will return the value of the available, prevailing evidence. The two parties may choose to opt for the same n or different n values but they usually diverge on the choice of the η value. The prosecutor's choice η_1 , placed at the numerator of a ratio, is always higher than the defenses' choice η_2 - placed at the denominator ($(\eta_1 - \eta_2) \geq 1$). For what matters in this paper, the classical LR formula [47] is therefore written as:

$$LR = \frac{Pr(E|Hp, n1, \eta1)}{Pr(E|Hd, n2, \eta2)} \quad (10)$$

Quantitative calculations

Background

Algebra has no power on quantitative calculations because all states of a quantitative deconvolution have unique probability values and algebraic simplifications cannot operate on these matrices. Leaving out algebra and staying with matrices is therefore our way to tackle the

quantitative issue. (The central role of matrices in addressing quantitative mixture analysis is in our view unsurprising, as this kind of analysis has been extensively addressed by adopting Bayesian networks, which are a tool enabling to harness complex matrices with comparatively little effort [48, 49]).

The issue of quantitative mixture analysis has been reviewed several times [50-54] and – according to most reviewers - a key issue in the field is finding a suitable way to infer the peak fraction values pertaining to each contributor from the whole of the mixture evidence. In a pioneer-paper, Cowell et al [55] used the gamma distribution to deduct all individual peak heights fractions that coalesce to shape the mixed STRs profile. Almost at the same time, Curran (2008; [23]) introduced Markov Chain Monte Carlo simulations to infer this distribution. These two methods have earned much momentum in the subsequent decade and today many dedicated software pieces are based on the gamma model or MCMC methods. Here, however, we will adopt a new, all empirical method for deducting the mixing ratio at a mixture.

An empirical method enabling to capture MR fractions

Deducting the ratio at which contributors mix a mixture is difficult because contributors tend to share alleles and to overlap within the evidence. This phenomenon is heavier when numerous contributors concur to produce the same trace and at binomial loci.

However, at certain states of a mixture deconvolution, a single contributor is able to give the whole of the reading signal pertaining to a single allele. At each of these non-overlapping genotypes (NOGs), the number of reads referring to the NOG, divided by the sum of all reads, returns the proportion at which the NOG-bearing contributor has entered in the mixture.

At SNPs there also exist some simple-overlapping genotype combinations (SOGs) enabling to deduct the individual contribution with additional but elementary computation. Binomial NOGs and SOGs combinations are shown in Fig 3, along with the corresponding formulas to adopt for

277 extracting the mixing ratio. When no POI is available, NOGs and SOGs are of no use. In this case,
 278 all genotypes will permute within each position in the matrix, each genotype will end up to
 279 belong to each contributor and all contributors will share the same quantity (0,5 to 0,5 at the
 280 2PM; 0,33: 0,3 : 0,33 at 3PMs; 0,25x4 at 4PMs etcetera). But when a NOG/ SOG bearing POI
 281 becomes available, NOGs/SOGs will no longer permute and the reconstruction of the POI
 282 quantity becomes possible. Unknown contributors will of course continue to permute and share
 283 identical fractiond of the remaining quantity. To make a numerical example: at a 3PM a POI
 284 amounting to 0,4 /1 will leave a 0,6 /1 to the two unknowns who will amount to 0,3 /1 each. When
 285 the POI is AA, NOGs and SOGs values generated from all AA (BB)n and AA (AB)n series will
 286 correspond to the same AA POI and they become equally possible values. Both values ar then
 287 considered for deriving the MR. A POI may be a true contributor or a false contributor. If true,
 288 his/her quantitative NOGs will cluster around just a definite value, net of measurements errors. If
 289 he is a false contributor, its quantitative measurements will flatten on the 1/n value. In turn, MRs
 290 extracted from true contributors will return high LR values, false contributors will have $LR \leq 1$.
 291 When true contributors have an MR proportion close to the 1/n value, they behave as if they
 292 were false contributors (LR falls below 1). If more the one POI is available, the respective NOGs
 293 can be extracted either separately or together. Because SNPs protocols are usually made of some
 294 hundreds loci, a significant number of NOGs and SOGs are usually intercepted at any given
 295 experiment.

296 Proportions deducted from NOGs and SOGs can be generalized to all loci (either overlapping or
 297 non-overlapping) and become a proxy value for the general mix ratio (MR).

298 **Fig 3. Non overlapping (NOCs) and simple overlapping (SOCs) combinations at SNPs.**

299 Genotypes to adopt for calculating the POI's mixing ratio are in red typeface; repetitive, non useful
 300 genotypes in black. Increasing the number of contributors leads to an elongation of the non-

301 useful genotypes whereas NOGs and SOGs kept outside the overlapping zone and remain unique.

302 Data and the formulas to adopt for calculating the MR are specified under the relevant scheme.

303 Introducing empirical MR values into probability calculations

304 Once deducted from the NOGs/SOGs statistics, the empirical MR become the 'expected' value

305 to retrieve at every single combination of genotypes entering in a likelihood matrix. 'Observed'

306 values of mixing ratios are in turn deducted from each, n^{th} deconvolution state of the mixture, by

307 simple algebraic operations reproducing those in figure 3 and placed within our worksheets (to be

308 downloaded at : www.vincepascali.it/docs/NITZq_SNP_2PM.xlsx;

309 www.vincepascali.it/docs/NITZq_SNP_3PM.xlsx; www.vincepascali.it/docs/NITZq_SNP_4PM.xlsx).

310 Observed and expected MR values are compared by setting series of chi square tests. The $(1-\chi^2)$

311 value deducted at each individual state will eventually return the probability of that state to exist

312 in the mixture.

313 Three indexes for each likelihood value calculation

314 Our three NITZq matrices distribute three statistical indexes: the $(1-\chi^2)$ MR values, the h index

315 (the minor read divided by the major read) and the population genotype frequency. These three

316 values embody as many independent properties of mixtures (the MR belongs to the specific

317 mixture; the h index is an autosomal property; the genotype frequency is from the population)

318 and they multiply at the level of each deconvolution state. All products are in turn summed to

319 give the value of any likelihood of interest.

320 Flow chart of NITZ calculations

321 Although our NITZq worksheets have no pretence to make an integrated software, each

322 flowcharts a whole series of 130-some SNPs Mendelian and quantitative data into a unique, semi-

323 automatic procedure that return perfectly reliable LR calculations.

324 The sequence of passages included within NITZq worksheet flow chart is as below.

325 **I. evidence input.** The appropriate worksheet of the NITZ library (NITZq 2PM, NITZq 3PM and
326 NITZq 4PM) is to be selected. Each worksheet has an “*input & MR*” spreadsheet that allows the
327 user to input the following data (from the worksheet’s top-left side onwards and in sequential
328 order per each locus): locus name; major read; minor read; gene frequency of the major read
329 gene; gene frequency of minor read gene;

330 **II. POI genotypes input;** all POI(s) genotypes have first to be code-classified as AA (homozygous
331 major read), BB (homozygous minor read), AB (major-minor heterozygous) according to the
332 conventional, major-read allele / minor-read allele code. Code-classified genotypes are then paste
333 into the input area as ‘reference data’ placed adjacent to the ‘evidence data’. This will trigger
334 hundred calculations into a series of cross-referenced worksheets.

335 **III. basic calculations:** specific algebraic functions placed in a “*calculations*” worksheet calculate
336 genotype frequencies, individual quantities and heterozygous ratio (h) at every record
337 corresponding to each state of the loci deconvolutions (presently NITZq allocates 133 loci).

338 **IV. calculation of the MR:** at the “*input & MR*” spreadsheet, NOGS/SOGs values are intercepted
339 based on the POI genotypes. Their values are summed and averaged and the quantity fraction
340 belonging to the POI is deducted. Finally, $(1 - Q_{POI})$ divided by the number of unknowns will give the
341 other fractions of MR. The array of the MR values is cross-referenced into thousands positions at
342 the ‘*calculations*’ worksheet.

343 **V. $(1 - \chi^2)$ calculation:** the individual quantities deducted at each deconvolution state are
344 compared to the NOGS/SOGs MR and transformed into $(1 - \chi^2)$ values (one value per record);

345 **VI likelihood calculations:** genotype frequencies, h values and $(1 - \chi^2)$ values are multiplied across
346 each deconvolution state and their products summed together to give the likelihood of choice

347 **VII calculation of the LR**

348 Likelihoods and LR calculations update instantly as soon as the evidence and POI input is pasted
349 into the “input & MR” spreadsheet . Computation time is virtually zero.

350 **VIII correction for incompatible cases**

351 Extremely imbalanced read pairs in the evidence may indicate that all the underlying mixture
352 contributors are homozygous and the evidence itself may qualify as single-allele ($Et\{a\}$). Because
353 no deconvolution and no quantity calculation applies to all tests bearing the $Et\{a\}$ evidence, these
354 cases need to be identified and brought back to algebra. To spot cases as these, we calculate the
355 h index at every locus evidence, set a threshold for the h-evidence value below which the locus is
356 declared ‘single-allele’ and reclassify. Spotting single-allele evidence is essential in order to identify
357 privileged Evidence/Reference assortments: $Et\{a\}$ vs POI_{AA} $Et\{a\}$ vs POI_{BB} $Et\{a\}$ vs POI_{AB} .
358 Generally, the first assortment tends to give high LRs placed well above 1, the latter two low/very
359 low LRs.

360 We indiscriminately apply this correction to the semi-continuous and the quantitative scheme.

361 **Testing datasets**

362 To test the performance of our formulas, matrices and worksheets and compare it with high
363 quality software we chose to download the “dataset 4: SNPs data” made available by Dr Ø. Bleka
364 at the open-source EUROFORMIX website (<http://www.EUROFORMIX.com/?q=data>). This dataset
365 contains evidentiary genotypic and quantitative data pertaining to several experimental 2PM
366 mixtures and 3PM mixtures designed to cover various mixing ratios under accurately controlled
367 conditions. ‘EUROFORMIX SNPs Data’ supplies typing results pertaining to a subset of 140 biallelic
368 SNPs [56]. It also includes three POI genotypic dataset corresponding to three contributors
369 (P_1, P_2, P_3) that have been actually used to build the evidence. An array of European gene
370 frequencies is also supplied. We singled out the XEN13_a (a two person mixture made of
371 individual P_1 and P_2 , mixing at five different MRs; 1:1; 1:3; 1:9; 3:1; 9:1) and the XEN46_a (a three

372 person mixture made with individual P1, P2 and P3, mixing at: 1:1:1; 5:1.1; 1:5:1; 1:1.5; 5:5:1;
373 5:1:5; 1:5:5) as datasets for our comparison test. After elimination of incomplete typing data, a
374 selection of 133 XEN13_a /XEN46 loci was eventually selected and brought to the statistical
375 analysis. In addition, we used data from forty random Swedish individuals (the much valued
376 courtesy of professor Andreas Tillmar, Department of Forensic Genetics and Forensic Toxicology,
377 National Board of Forensic Medicine, Linköping, Sweden) as non-contributor datasets.
378 Each group of data drawn from this collection was thoroughly analysed by adopting the 2-person,
379 3-person and 4 person mixing scheme irrespective of the fact that the original evidence would
380 originate from a two-person or a three person's . All data were independently processed by using
381 the EUROFORMIX shareware (version 3.0.3) and then by the collection of all NITZq worksheets.

382 Results

383 MRs from XEN 13a / XEN 46a, three true contributors as POI and a 384 variable amount of unknowns

385 Fig 4 and Fig 5 show twelve series of mix ratio predictions as they come out from our NITZq
386 2PM, 3PM and 4 PM worksheets and from EUROFORMIX. The evidentiary datasets were an
387 experimental three person mixture (XEN46a) containing three contributors P1,P3,P2 mixing at
388 seven different ratios (1:1:1; 5:5:1; 5:1:5; 1:5:5; 5:1:1; 1:1:5; 1:5:1; quoted in the order
389 P1,P3;P2) and a two-person mixture (Xen13a) containing two contributors P1 and P2 blending at
390 five different balancing ratios (1:1;1;13;1:9; 3:1; 9:1; first quoted quantity is P1). P1, P3 and P2
391 were in turn individually used as person of interest along with one (KU), two (KUU) and three
392 (KUUU) additional unknown contributors (KU; KUU; KUUU). The POI is here always a 'true'
393 known contributor— i.e. a person whose DNA had been actually used to cast the mixture evidence
394 - its genotypic array learned from the EUROFORMIX website. Just one of the three hypotheses on

the number of contributors used to interrogate each evidentiary example was ‘true’ based on what’s declared in the EUROFORMIX website. The other predictions were assuming a wrong number of contributors. These data shows that NITZq quantitative POI predictions compare well to those issued by the EUROFOMIX software whereas predictions on unknown contributors operated by the two methods may occasionally diverge: in principle, NITZ always flattens the unknown MR proportions to exactly the same value; EUROFORMIX occasionally assigns differential values, other times distributes just one value to all unknowns. Predicting quantities under a wrong assumption about the true number of contributors has a limited effect on the effect on the POI’s share. In principle, expanding the contributors’ number shrinks the quantitative proportion left to a true POI.

Fig 4. Amount of DNA assigned to a true contributor POI at a three-person-based experimental evidence (XEN46a)

Fig.5. Amount of DNA assigned to a true contributor POI at a two-person-based experimental evidence (XEN13a)

MRs, with false contributors in the role of POI

Another series of MR predictions were produced by replacing the ‘true’ P1,P3, P2 with false contributors in the role of POI. False contributors were taken one by one among a population of forty unrelated Swedes (A. Tillman’s data). Predictions followed the same scheme as before but for simplicity each POI was matched to a fewer evidentiary datasets (XEN13, ratios: 1:1; 1:9; XEN46a ratios: 1:1:1; 1:5:5; 1:1:5). With both NITZ and EUROFORMIX, quantitative fractions predicted to each of the forty POIs were so close one to another that they could be conveniently averaged with little standard deviation, as shown in Fig 6 and Fig 7 . Figures here reflect the different approach of the two methods: EUROFORMIX seeks for the best fitting quantitative proportions in hundred random simulations, NITZ singles out values tending to $1/n$, which is the

result of a free deconvolution. We envision the $1/n$ value is eventually reached by predicting MRs over an infinite number of loci. Increasing the number of contributors has the effect of halving the quantitative fraction assigned by EUROFORMIX. NITZq moves in pretty much the same direction, but the decrease in the POI quantitative proportion has its floor in $1/n$. Matching the same POIs to balanced or unbalanced mixtures has little effect on the predicted proportions.

Fig 6. Average amount of DNA assigned to a cohort of 40 false contributors placed (one at the time) in the role of POI at a three-person-based experimental evidence (XEN46a)

Fig 7. Amount of DNA assigned to a cohort of 40 false contributors placed (one at the time) in the role of POI at a two-person-based experimental evidence (XEN13a)

LR values from true contributors as POI

Log₁₀(LR) values calculated from the two tutorial datasets are in Fig 8 and Fig 9. To produce this data, we set NITZq to call the biallelic evidence every time the minor allele read would be $\geq$ to the major read multiplied by 0,05 and call the mono-allelic below this value. EUROFORMIX calculations were based on an absolute minor allele calling threshold of 5 reads and a $P(D)=0.05$. Semi-continuous LRs are of comparable value if the two systems of calculation end up to call the same homozygous genotypes within the whole locus evidence. Betting on the wrong number of contributors does not significantly affect the semi-continuous LRs of either system. Both NITZq and EUROFORMIX return comfortably high values of quantitative LRs at tests where POIs match imbalanced evidence. However, NITZ underestimates LRs by 1 to sometimes 10 orders of magnitude (or bans), compared to EUROFORMIX, when the POI predominates in quantity within the evidence. To investigate about where such a difference can come from, we compared individual locus specific LRs obtained from selected mixture schemes to the corresponding single –source LR values ($1/P(GT)$, or 1 divided by the POI's genotype frequency). We consequently noticed that EUROFORMIX ver. 3.3. often exceeds the $1/P(GT)$ value, whereas

NITZq never behaves like this. To cite an example, at a typical P1UU/UUU calculation based on XEN46a tutorial dataset ratio 5:1:1 (with P1= POI and P1 DNA quantity exceeding five times P3 and P2 in the mixture content) this happened at more than fifty over 133 overall loci (data not shown). According to Cowell et al [27] the weight of evidence (WOE) to draw from a suspect's profile compared to a matching mixture cannot be stronger than that obtained when the suspect matches the single-source stain, or :

$$LR = \frac{Pr(Et|Hp)}{Pr(Et|Hd)} \leq 1/Pr(GT)$$

(where Pr(GT) is the suspect's genotype frequency. We believe that this scores a strong point on NITZ's side and earns credibility to its predictions.

Fig 8. Log10(LRs) pertaining to three true contributors matched (one by one) as POI to an experimental three person mixture (XEN46a). EUROFORMIX and NITZq calculations are shown in comparison

Fig 9. Log10(LRs) pertaining to three true contributors matched (one by one) as POI to an experimental two person mixture (XEN13a) . EUROFORMIX and NITZq calculations are shown in comparison

LR values for false contributors as POI

These datasets were produced under the same scheme of analysis detailed in the previous chapter. An excerpt of the whole data showing the maximum and minimum values of LR, along with the internal percentile distribution is in Fig 10 and Fig 11, where NITZ and EUROFORMIX are shown in direct comparison (log10(LR) are shown). It is easily noticed that NITZq returns much lower LR on false contributors, thus ruling out type-two errors (false inclusions) more efficiency than EUROFORMIX. This is essentially on account of three issues:

- a. NITZ is highly efficient in spotting the homozygous evidence, thanks to the dynamic threshold (based on the minor vs major read ratio) used to call the minor read allele;

- b. As a consequence, NITZ allows for more exclusion cases to emerge - with evidence $E\{a\}$ either facing a $POI=AB$ or a $POI=BB$;
- c. NITZ deals with these two classes of exclusion by algebra, according to equ. 6-9; this results in low ($E\{a\}$ vs $K=AB$) or very low ($E\{a\}$ vs $K=BB$) LR values.

Fig 10. Log10(LRs) pertaining to forty false contributors matched (one by one) as POI to an experimental three person mixture (XEN46a). EUROFORMIX and NITZq calculations are shown in comparison

Fig 11. Log10(LRs) pertaining to forty false contributors matched (one by one) as POI to an experimental two person mixture (XEN13a). EUROFORMIX and NITZq calculations are shown in comparison

Correlation between MR and LR values, with true contributors as POI

The LR/MR relationship was investigated at both the XEN46a and XEN13a evidence, by matching P1, P2 P3 (the 'true' contributors) as POI to variously imbalanced mixtures. For simplicity, LR values were calculated by adopting the true number of contributors (at XEN13a: KU/UU; at XEN46a: KUU/UUU). Then the $\text{Log}_{10}(\text{LR}_{\text{quant}})$ minus the $\text{Log}_{10}(\text{LR}_{\text{semicont}})$ value was paired to the quantitative fraction predicted to each POI. Pairs of values were plotted together, regardless of the difference in genetic profile originating the data-points.

We found that when true contributors are chosen as POI, a direct proportional relationship between the POI MR proportion and the $\text{LOG}_{10}(\text{LR})$ value emerges from the NITZ-generated series of data (Fig 12; Fig 13). We concluded that the higher MR fraction of the POI the higher LR value. This relationship holds well even when a wrong number of contributors is used to calculate the LRs (data not shown). Data generated by EUROFORMIX behaved in the same way (data not shown).

Fig 12. Plots of POI's DNA amount vs $\text{Log}_{10}(\text{LR})$ value in a series of true contributors tests.

489 The evidence is XEN46_a, POIs are P1,P2,P3 (all downloaded from EUROFORMIX website). The
490 NITZq 3PM worksheet was used to produce these results.

491 **Fig 13. Plots of POI's DNA amount vs Log10(LR) value in a series of true contributors tests.**

492 The evidence is XEN13_a, POIs are P1,P2 (all downloaded from EUROFORMIX website). The NITZq
493 2PM worksheet was used to produce these results.

494 **MR and LR relationship – with false contributors as POI**

495 This relationship was investigated by the same scheme of analysis as adopted in the previous
496 chapter. Forty false contributors were chosen – one at the time - as the POI in the following
497 schemes of analyses: XEN46a ratio 1.1:1, KUU/UUU; XEN13a ratio 1.1, KU/UU. At false
498 contributors tests, no coherent relationship was found among MR and le LR (Fig 14; Fig 15).

499 **Fig 14. Plots of POI's DNA amount vs Log10(LR) value in a series of false contributors tests.**

500 The evidence is XEN46_a ratio 1:1:1, POIs are forty unrelated individuals. The NITZq 3PM
501 worksheet was used to produce these results.

502 **Fig 15. Plots of POI's DNA amount vs Log10(LR) value in a series of false contributors tests.**

503 The evidence is XEN13_a ratio 1:1, POIs are forty unrelated individuals. The NITZq 2PM worksheet
504 was used to produce these results.

505 **Discussion**

506 We have introduced a series of original procedures meant to re-define the way semi-continuous
507 and quantitative calculations are carried out at binomial markers.

508 On the **semi-continuous** side of the issue, we vindicate the viewpoint that the routine binomial
509 evidence at a mixture cannot be but bi-allelic ($E\{a;b\}$) and that this predominant example of SNPs
510 evidence is quite well modelled by matrices of multiple, equally possible genotypic permutations.
511 Within these discrete framework of probability matrices, all genotypes entering in the same

512 permutation state multiply together and all permutation states products sum up, so as to give the
513 true value of a likelihood. As long as calculations rest on only genotype probabilities, the matrix
514 scheme of calculation can be easily reduced to algebra - and we have here introduced a series of
515 notations that make previously reported formulas [34] of universal use in the semi-continuous
516 context. Interestingly, matrix-based calculations and algebraic calculations give identical results,
517 thus implying that the two underlying methods (matrices; algebra) are sound and each is
518 consistent to the other. If binomial matrices and algebraic formulas are used, semi-continuous
519 inference at SNPs is reduced to the application of simple arithmetic and it does not require
520 precompiled software. This way to calculate does not undergo any limitation in the number of
521 contributors and in the number of POIs to compute on- as often dedicated software does.
522 In turn, we interpret mono-allelic evidence at a SNPs mixture ($Et\{a\}$) as a ‘residual’ evidence – the
523 product of a biallelic evidence having undergone dropout events. We assign the $Et\{a\}$ locus
524 evidence only when the h index at a given locus falls below a predetermined, dynamic threshold
525 reflecting extreme imbalance among the minor and the major read of an individual NGS record.
526 Albeit residual in their practical occurrence, examples of the monoallelic evidence $Et\{a\}$ emerging
527 at a SNPs protocol can modify the inferential value of an identity test to a big extent, by either
528 considerably increasing the LR (case: $Et\{a\}/POI_{AA}$) or considerably lowering it (cases: $Et\{a\}/POI_{BB}$
529 ; $Et\{a\}/POI_{BB}$). To calculate the relevant LRs at these ‘residual but important’ cases we have
530 worked out nine algebraic solutions covering all possible instances of $Et\{a\}/POI$ genotype
531 assortment. Within these formulas the ‘ h ’ index – or the minor vs major read ratio of a typical
532 NGS experiment – plays the role of an acceptable proxy of the dropout probability, and no
533 experimental $P(DO)$ values are needed. The whole procedure just here summarized becomes, in
534 our view, an acceptable equivalent of using the split-drop approach.

535 As far as the **quantitative** calculation is concerned, we here introduce an empirical way to capture
536 the ratio at which any give POI of interest mix with the other unknown mixture contributors. We
537 examine all permutations containing the POI- genotype at all loci of any standard testing
538 protocol of choice and, among these, we discern/extract all those that do not involve allele
539 overlaps between the POI's genotype and the unknowns' (NOCs) - along with all further cases
540 involving very simple POI-Unknowns overlaps (SOCs). By simple arithmetic we extract the
541 empirical POI share at all of these privileged loci, we average all values and we determine an
542 expected 'POI-vs-Unknowns' ratio. Compliance between this unique expected values and all
543 observed MR values pertaining to all existing deconvolution states (measured as a $(1-\chi^2)$ value)
544 gives us the probability of any permutations states to exist on the quantitative side.
545 Compared to other methods [23-55] our method for calculating the mixture MR is surprisingly
546 simple. Its computation within a spreadsheet is practically instantaneous and it ends up to give
547 back quantitative ratios that approximate well the MCMC and MLE estimates.
548 Unlike other methods, our MRs flatten the Unknowns share to the same figure, due to the
549 phenomenon of quantity permutations across the matrix. NOCs-SOCs based $(1-\chi^2)$ values are
550 true probability values and they are independent of the genotype probabilities and the h indexes.
551 These three independent probabilities are therefore multiplied together to yield the value of a
552 likelihood, according to a unique procedure of our conception ('NITZq' procedure).
553 We extensively tested NITZq by choosing a 133 c SNPs loci protocol [actually, a subset of the 140
554 loci panel of Grandell et al 2016], by using the "dataset 4: SNPs data" package available at the
555 EUROFORMIX website as evidence and by assuming the relevant EUROFORMIX calculations as the
556 benchmark. We here show that when true contributors are tested as POI, NITZq computations
557 return LR values of comparable magnitude as those returned by EUROFORMIX. Type-one errors
558 (mediocre or negative LOGLRs) occasionally occur with NITZ – essentially when the true

contributor's quantitative share tends to disappear from the mixture. Additionally, when the POI is a true contributor and is quantitatively predominant, NITZ usually gives back a lower value of the evidence (WOE) than EUROFORMIX. As an explanation of this latter phenomenon, we report that the EUROFORMIX individual LR's often exceed the corresponding single-source WOE (namely: one over the relevant genotype frequency value) – usually regarded as a ceiling value not to be crossed while evaluating the presence of individual profiles within a mixture. NITZq never oversteps the relevant single-source WOE and it therefore qualifies as a fully realistic approach to the quantitative analysis. When false contributors are chosen as POI, a much lower LR is gathered by NITZq, compared to EUROFORMIX. Type-two errors (false inclusions) occasionally occur with EUROFORMIX whereas they practically do not occur with NITZq. We believe that this is due to the circumstance that NITZ spots incompatible events much more efficiently than EUROFORMIX. When a true contributor is a POI, a linear relationship exists among its quantitative share in the mixture and the $\log_{10}(\text{LR})$ value, as calculated by NITZ. When the POI is a false contributor, this relationship becomes incoherent and graphically sparse. With NITZ, betting on the wrong number of contributors does not necessarily result in lower LR's. So the rule of three invoked by Bleka et al [36] has limited application with NITZ.

References

1. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526: 68–74.
2. Sobrino S, Carracedo A. SNP typing in forensic genetics: a review. *Methods Mol. Biol.* 2005; 297: 107-26.
3. Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, et al. SNPforID Consortium, Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel, *Forensic Sci Int Genet.* 2007; 1(2): 180-185.

- 583 4. Johansen P, Andersen JD, Børsting C, Morling N. Evaluation of the iPLEX® Sample ID Plus Panel
584 designed for the Sequenom MassARRAY® system. A SNP typing assay developed for human
585 identification and sample tracking based on the SNPforID panel. Forensic Sci Int Genet. 2013; 7(5):
586 482-487.
- 587 5. Churchill JD, Schmedes SE, King JL, Budowle B, Evaluation of the Illumina (®) beta version
588 ForenSeq™ DNA signature prep kit for use in genetic profiling, Forensic Sci Int Genet. 2016; 20:
589 20–29.
- 590 6. Zhang S, Bian Y, Zhang Z, Zheng H, Wang Z, Zha L, et al. Parallel analysis of 124 universal SNPs
591 for human identification by targeted semiconductor sequencing. Sci. Rep. 2015; 5: 18683.
- 592 7. Guo F, Zhou Y, Song H, Zhao J, Shen H, Zhao B, Liu F, Jiang X. Next generation sequencing of
593 SNPs using the HID-Ion AmpliSeq™ Identity Panel on the Ion Torrent PGM™ platform, Forensic Sci.
594 Int. Genet. 25 (2016) 73-84.
- 595 8. M. de la Puente M, C. Phillips C, C. Santos C, Fondevila CM, Carracedo A, Lareu MV, Evaluation
596 of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing. Forensic
597 Sci Int Genet. 2017; 28: 35-43.
- 598 9. Guo F, Yu J, Zhang L, Li J, Massively parallel sequencing of forensic STRs and SNPs using the
599 illumina® ForenSeq™ DNA signature prep kit on the MiSeq FGx™ forensic genomics system,
600 Forensic Sci Int Genet. 2017; 31: 135–148.
- 601 10. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, et al. Developmental validation of
602 the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic
603 DNA Casework and Database Laboratories, Forensic Sci Int Genet. 2017; 28: 52–70.
- 604 11. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit – MPS
605 forensic application for the MiSeq FGx™ benchtop sequencer, Forensic Sci Int Genet. 2017; 28 :
606 188–194.

- 607 12. Li LM, Wang Y, Yang SP, Xia MY, Yang YJ, Wang JC, et al. . Genome-wide screening for highly
608 discriminative SNPs for personal identification and their assessment in world populations, Forensic
609 Sci Int Genet. 2017; 28: 118–127.
- 610 13. Hussing C, Huber C, Bytyci CR, Mogensen HS, Morling N, Børsting NC, Sequencing of 231
611 forensic genetic markers using the MiSeq FGx™ forensic genomics system – an evaluation of the
612 assay and software, Forensic Sci Res. 2018; 3: 111–123.
- 613 14. Hwa HL, Chung WC, Chen PL, Lin CP, Li HY, Yin HI, Lee JCI. A 1204- single nucleotide
614 polymorphism and insertion–deletion polymorphism panel for massively parallel sequencing
615 analysis of DNA mixtures, Forensic Sci Int Genet. 32 (2017) 94–101.
- 616 15. Yousefi S, Abbassi-Dalooi T, Kraaijenbrink T, Vermaat M, Mei H, van 't Hof P, et al. A SNP panel
617 for identification of DNA and RNA specimens. BMC Genomics. 2018; 19(1): 90.
- 618 16. Erwen H, Changhui HL, Jingjing Z, Xiaolong H, Weian HD, Yuanjian H, et al. Genome-wide screen
619 for universal individual identification SNPs based on the HapMap and 1000 Genomes databases,
620 Nature Scientific Reports 2018; 8:5553 | DOI:10.1038/s41598-018-23888
- 621 17. Ricke DO, Shcherbina A, Michaleas A, Fremont-Smith P, Grigora MS. SNPs: Optimized Analysis
622 of SNPs for DNA Forensics, J Forensic Sci, November 2018 , Vol. 63, No. 6
- 623 18. Avent I, Kinnane AG, Jones N, Petermann I, Daniel R, Gahan ME, McNevin D, The QIAGEN 140-
624 locus single-nucleotide polymorphism (SNP) panel for forensic identification using massively parallel
625 sequencing (MPS): an evaluation and a direct-to-PCR trial, Int J Legal Med. 2019; 133: 677–688
- 626 19. C. Phillips C, Manzo L, de la Puente M, Fondevila M, Lareu MV. The MASTiFF panel—a versatile
627 multiple-allele SNP test for forensics. 134:441–450.
- 628 20. Petrovick MS, Boettcher T, Fremont-Smith P, Peragallo C, Darrell O. Ricke DO, Watkins J,
629 Schwoebel J. Analysis of complex DNA mixtures using massively parallel sequencing of SNPs with
630 low minor allele frequencies. Forensic Sci Int Genetics. 2020; 46: 102234.

- 631 21. Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J, Interpreting DNA Mixtures, J
632 Forensic Sci. 1997; 42: 213-222.
- 633 22. Evett IW, Gill P, Lambert JA, Taking account of peak areas when interpreting mixed DNA profiles,
634 J Forensic Sci. 1998; 43:62-69.
- 635 23. Curran JM, A MCMC method for resolving two person mixtures, Science and Justice. 2008; 48:
636 168-177.
- 637 24. Wang T, Xue N, Douglas Birdwell J, Least-Square deconvolution: a framework for interpreting
638 short tandem repeat mixtures, J Forensic Sci. 2006; 51:1284-1297.
- 639 25. Puch-Solis R, Clayton T. Evidential evaluation of DNA profiles using a discrete statistical model
640 implemented in the DNA LiRa software, Forensic Sci Int Genet. 2014; 11: 220-228.
- 641 26. Cowell R, Lauritzen SL, Mortera J. A gamma model for DNA mixture analyses, Bayesian Analysis.
642 2010; 2: 333-348.
- 643 27. Cowell RG, Graversen T, Lauritzen SL, Mortera J, Analysis of forensic DNA mixtures with
644 artefacts, Applied Statistics. 2015; 64:1-32.
- 645 28. Phillips C, Prieto L, Fondevila M, Salas A, Gómez-Tato A, Álvarez-Dios J, et al. Ancestry analysis
646 in the 11-M Madrid bomb attack investigation, PLoS One 2009; 4: e6583.
- 647 29. Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, et al. Resolving individuals
648 contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping
649 microarrays, PLoS Genet. 2008; 4: e1000167.
- 650 30. Voskoboinik L, Darvasi A, Forensic identification of an individual in complex DNA mixtures,
651 Forensic Sci. Int. Genet. 2011; 5: 428-35.
- 652 31. Isaacson J, Schwoebel E, Shcherbina A, Ricke D, Harper J, Petrovick M, et al. Robust detection
653 of individual forensic profiles in DNA mixtures, Forensic Sci. Int. Genet. 14 (2015) 31-37.

- 654 32. Clayton D, On inferring presence of an individual in a mixture: a Bayesian approach. *Biostatistics*.
655 2010; 11(4): 661-73.
- 656 33. Egeland T, Fonneløp AE, Berg PR, Kent M, Lien S, Complex mixtures: a critical examination of a
657 paper by Homer et al. *Forensic Sci Int Genet*. 2012; 6(1): 64-69.
- 658 34. Voskoboinik L, Ayers SB, LeFebvre AK, Darvasi A, SNP-microarrays can accurately identify the
659 presence of an individual in complex forensic DNA mixtures, *Forensic Sci. Int. Genet*. 2015; 16:
660 208–215.
- 661 35. Gill P, Haned H, Eduardoff M, Santos C, Phillips C, Parson W, The open-source software LRmix
662 can be used to analyse SNP mixtures, *Forensic Sci Int Genet. Supplement Series* 2015; 5: e50–e51
- 663 36. Bleka Ø, Eduardoff M, Santos C, Phillips C, Parson W, Gill P, Open source software
664 EUROFORMIX can be used to analyse complex SNP mixtures, *Forensic Sci Int Genet*. 2017; 31:
665 105–110.
- 666 37. Lynch PC, Cotton RW. Determination of the possible number of genotypes which can contribute
667 to DNA mixtures: non computer assisted deconvolution should not be attempted for greater than
668 three contributors. *Forensic Sci Int Genet*. 2018; 37: 235-240.
- 669 38. Benshop CG, Haned H, Jeurissen L, Gill P, Sijen T, The effect of varying the number of contributors
670 on likelihood ratios for complex DNA mixtures, *Forensic Sci. Int. Genet*. 19 (2015) 92-99
- 671 39. Tvedebrink T, Eriksen PS, Mogensen HS, Morling N, Estimating the probability of allelic drop-out
672 of STR alleles in forensic genetics, *Forensic Sci Int Genet*. 2009; 3: 222–226.
- 673 40. Buckleton JS, Kelly H, Bright JA, Taylor D, Tvedebrink T, Curran JM, Utilising allelic dropout
674 probabilities estimated by logistic regression in casework, *Forensic Sci Int Genet*. 2014; 9: 9–11.
- 675 41. Haned H, Slooten K, Gill P, Exploratory data analysis for the interpretation of low template DNA
676 mixtures. *Forensic Sci Int Genet*. 2012; 6(6): 762-74.

677 42. Gill P, Haned H, Eduardoff M, Santos C, Phillips C, Parson W. The open-source software LRMix
678 can be used to analyse SNP mixtures. *Forensic Sci Int Genetics*. 2015; 5: e50-e51

679 43. Sharma V, Chow HY, Siegel D, Wurmbach E. Qualitative and quantitative assessment of
680 Illumina's forensic STR and SNP kits on MiSeq FGx™. *PONE*. 2017; 12(11): e0187932.

681 44. Kelly H, Bright JA, Curran LM, Buckleton J, Modelling heterozygote balance in forensic DNA
682 profiles. *Forensic Sci Int Genetics*. 2012; 6: 729-734.

683 45. R.G. Cowell RG. Computation of marginal distributions of peak-heights in electropherograms for
684 analysing single source and mixture STR DNA samples, *Forensic Sci Int Genet*. 2018; 35: 164–168.

685 46. Gill P, Gusmão L, Haned H, Mayr WR, Morling N, Parson W, et al. DNA commission of the
686 International Society of Forensic Genetics: recommendations on the evaluation of STR typing results
687 that may include drop-out and/or drop-in using probabilistic methods. *Forensic Sci Int Genet*. 2012;
688 6: 679–688.

689 47. Lindley DV. A problem in forensic science. *Biometrika*. 1977; 64: 207-213.

690 48. Dawid AP, Mortera J, Pascali VL, van Boxel V. Probabilistic expert system for forensic inference
691 from genetic markers *Scand J Stat*. 2002 29: 577-595.

692 49. Mortera J, Dawid AP, Lauritzen SL. Probabilistic expert systems for DNA mixture profiling.
693 *Theor Popul Biol*. 2003; 63(3):191-205.

694 50. Mortera J. DNA mixtures in forensic investigations: the statistical state of the art. *Annu Rev*
695 *Stat appl*. 2020; 7:111-142.

696 51: Yang J, Lin D, Deng C, Li Z, Pu Y, Yu Y, et al. The advances in DNA mixture interpretation.
697 *Forensic Sci Int* 2019; 301:101-106.

698 52: Oldoni F, Podini D. Forensic molecular biomarkers for mixture analysis. *Forensic Sci Int Genet*.
699 2019; 41:107-119.

700 53: Coble MD, Bright JA. Probabilistic genotyping software: An overview. Forensic Sci Int Genet.
701 2019; 38:219-224.

702 54: Tao R, Wang S, Zhang J, Zhang J, Yang Z, Sheng X, Hou Y, Zhang S, Li C. Separation/extraction,
703 detection, and interpretation of DNA mixtures in forensic science (review). Int J Legal Med. 2018
704 132(5):1247-1261.

705 55. Cowell, R, Lauritzen SL and Mortera J. A gamma model for DNA mixture analysis. Baysn Anal.
706 2007; 2: 333-348.

707 56. Grandell I, Samara R, Tillmar AO. A SNPs panel for identity and kinship testing using massive
708 parallel sequencing Int J legal med 2016; 130 : 905-914.

		evidence $\rightarrow$		
		Et{a}	Et{b}	Et{a,b}
POIs $\downarrow$	kt{a}	$Kt = Et$	$Kt \cap Et = \emptyset$	$Kt \subset Et$
	kt{b}	$Kt \cap Et = \emptyset$	$Kt = Et$	$Kt \subset Et$
	kt{a,b}	$Et \subset Kt$	$Et \subset Kt$	$Kt = Et$
	kt{$\emptyset$}	-	-	Deconv.

Figure

evidence →

POIs ↓

	$\text{Et}\{a\} \text{ aut } (\text{Et}\{a, b_d\})$	$\text{Et}\{b\} \text{ aut } \text{Et}\{a_d, b\}$	$\text{Et}\{a, b\}$
$\text{kt}\{a\}$	$(1-h) * a^{2(n-\eta)}$	$h^2 * a^{2(n-\eta)}$	$1 - a^{2(n-\eta)}$
$\text{kt}\{b\}$	$h^2 * b^{2(n-\eta)}$	$(1-h) * b^{2(n-\eta)}$	$1 - b^{2(n-\eta)}$
$\text{kt}\{a, b\}$	h	h	1
$\text{kt}\{\emptyset\}$			$1 - a^{2n} - b^{2n}$

Figure

	NOG	repetitive genotypes			
2PM	AA	BB			
3PM	AA	BB	BB		
4PM	AA	BB	BB	BB	
5PM	AA	BB	BB	BB	BB

data:

r_a reads of allele a

r_b reads of allele b

$r_a > r_b$

POI's genotype : AA

POI's quantitative fraction = $r_a / (r_a + r_b)$

	NOG	repetitive genotypes			
2PM	BB	AA			
3PM	BB	AA	AA		
4PM	BB	AA	AA	AA	
5PM	BB	AA	AA	AA	AA

data:

r_a reads of allele a

r_b reads of allele b

$r_a > r_b$

POI's genotype : BB

POI's quantitative fraction = $r_b / (r_a + r_b)$

	SOG	repetitive genotypes			
2PM	AA	AB			
3PM	AA	AB	AB		
4PM	AA	AB	AB	AB	
5PM	AA	AB	AB	AB	AB

data:

r_a reads of allele a

r_b reads of allele b

POI's genotype : AA

POI's quantitative fraction = $(r_a \cdot r_b) / (r_a + r_b)$

	SOG	repetitive genotypes			
2PM	AB	AA			
3PM	AB	AA	AA		
4PM	AB	AA	AA	AA	
5PM	AB	AA	AA	AA	AA

data:

r_a reads of allele a

r_b reads of allele b

POI's genotype : AB

POI's quantitative fraction = $2 \cdot r_b / (r_a + r_b)$

Xen46a (evidence)	selected	Euroformix 2PM		NITZq 2PM		Euroformix 3PM			NITZq 3PM			Euroformix 4PM				NITZq 4PM			
true content	person of interest	DNA amount		DNA amount		DNA amount			DNA amount			DNA amount				DNA amount			
(real contributors and ratios)	(POI)	POI	U	POI	U	POI	U	U	POI	U	U	POI	U	U	U	POI	U	U	U
P1=1;P3=1;P2=1	P1	0,59	0,41	0,64	0,36	0,49	0,26	0,26	0,43	0,29	0,29	0,45	0,18	0,18	0,18	0,33	0,22	0,22	0,22
	P3	0,56	0,44	0,63	0,37	0,46	0,27	0,27	0,41	0,30	0,30	0,42	0,19	0,19	0,19	0,31	0,23	0,23	0,23
	P2	0,57	0,43	0,58	0,42	0,30	0,35	0,35	0,37	0,31	0,31	0,24	0,25	0,25	0,25	0,28	0,24	0,24	0,24
P1=5;P3=5;P2=1	P1	0,65	0,35	0,68	0,32	0,55	0,22	0,22	0,46	0,27	0,27	0,52	0,16	0,16	0,16	0,36	0,21	0,21	0,21
	P3	0,57	0,43	0,66	0,34	0,51	0,25	0,25	0,43	0,28	0,28	0,47	0,18	0,18	0,18	0,34	0,22	0,22	0,22
	P2	0,44	0,56	0,52	0,48	0,16	0,42	0,42	0,31	0,34	0,34	0,10	0,30	0,30	0,30	0,23	0,26	0,26	0,26
P1=5;P3=1;P2=5	P1	0,71	0,29	0,76	0,24	0,64	0,18	0,18	0,57	0,21	0,21	0,61	0,13	0,13	0,13	0,47	0,18	0,18	0,18
	P3	0,37	0,63	0,52	0,48	0,22	0,39	0,39	0,30	0,35	0,35	0,16	0,28	0,28	0,28	0,22	0,26	0,26	0,26
	P2	0,49	0,51	0,58	0,42	0,35	0,33	0,33	0,36	0,32	0,32	0,29	0,24	0,24	0,24	0,27	0,24	0,24	0,24
P1=1;P3=5;P2=5	P1	0,39	0,61	0,54	0,46	0,24	0,38	0,38	0,32	0,34	0,34	0,18	0,27	0,27	0,27	0,24	0,25	0,25	0,25
	P3	0,68	0,32	0,72	0,28	0,60	0,20	0,20	0,51	0,24	0,24	0,57	0,14	0,14	0,14	0,41	0,20	0,20	0,20
	P2	0,48	0,52	0,59	0,41	0,34	0,33	0,33	0,37	0,32	0,32	0,29	0,24	0,24	0,24	0,28	0,24	0,24	0,24
P1=1;P3=1;P2=5	P1	0,80	0,20	0,86	0,14	0,76	0,12	0,12	0,72	0,14	0,14	0,74	0,09	0,09	0,09	0,63	0,12	0,12	0,12
	P3	0,40	0,60	0,49	0,51	0,25	0,56	0,19	0,26	0,37	0,37	0,20	0,12	0,12	0,56	0,19	0,27	0,27	0,27
	P2	0,34	0,66	0,52	0,48	0,20	0,59	0,21	0,31	0,35	0,35	0,16	0,58	0,13	0,13	0,23	0,26	0,26	0,26
P1=5;P3=1;P2=1	P1	0,48	0,52	0,59	0,41	0,34	0,33	0,33	0,37	0,31	0,31	0,29	0,24	0,24	0,24	0,28	0,24	0,24	0,24
	P3	0,44	0,56	0,56	0,44	0,31	0,34	0,34	0,35	0,33	0,33	0,26	0,25	0,25	0,25	0,26	0,25	0,25	0,25
	P2	0,65	0,35	0,69	0,31	0,56	0,22	0,22	0,49	0,25	0,25	0,53	0,16	0,16	0,16	0,39	0,20	0,20	0,20
P1=1;P3=5;P2=1	P1	0,38	0,62	0,53	0,47	0,27	0,43	0,29	0,29	0,35	0,35	0,23	0,44	0,16	0,16	0,22	0,26	0,26	0,26
	P3	0,59	0,41	0,83	0,17	0,73	0,14	0,14	0,67	0,16	0,16	0,71	0,10	0,10	0,10	0,58	0,14	0,14	0,14
	P2	0,36	0,64	0,50	0,50	0,18	0,47	0,35	0,28	0,36	0,36	0,12	0,46	0,21	0,21	0,21	0,26	0,26	0,26

Figure

Xen13a (evidence) true content	selected person of interest	Euroformix 2PM DNA amount		NITZq 2PM DNA amount		Euroformix 3PM DNA amount			NITZq 3PM DNA amount			Euroformix 4PM DNA amount				NITZq 4PM DNA amount			
(real contributors and ratios)	(POI)	POI	U	POI	U	POI	U	U	POI	U	U	POI	U	U	U	POI	U	U	U
P1=1;P2=1	P1	0,70	0,30	0,77	0,23	0,66	0,17	0,17	0,58	0,21	0,21	0,66	0,17	0,17	0,00	0,48	0,17	0,17	0,17
	P2	0,49	0,51	0,62	0,38	0,45	0,53	0,03	0,40	0,30	0,30	0,45	0,01	0,01	0,53	0,31	0,23	0,23	0,23
P1=1;P2=3	P1	0,48	0,52	0,66	0,34	0,41	0,01	0,58	0,44	0,28	0,28	0,41	0,00	0,58	0,00	0,34	0,22	0,22	0,22
	P2	0,67	0,33	0,73	0,27	0,64	0,01	0,35	0,54	0,23	0,23	0,64	0,01	0,35	0,01	0,44	0,19	0,19	0,19
P1=1;P2=9	P1	0,30	0,70	0,54	0,46	0,20	0,79	0,01	0,31	0,35	0,35	0,20	0,01	0,01	0,79	0,23	0,26	0,26	0,26
	P2	0,84	0,16	0,88	0,12	0,82	0,13	0,06	0,76	0,12	0,12	0,82	0,13	0,05	0,00	0,67	0,11	0,11	0,11
P1=3;P2=1	P1	0,86	0,14	0,90	0,10	0,85	0,08	0,08	0,79	0,10	0,10	0,84	0,08	0,00	0,08	0,72	0,09	0,09	0,09
	P2	0,21	0,79	0,52	0,48	0,17	0,83	0,01	0,29	0,35	0,35	0,17	0,00	0,83	0,00	0,22	0,26	0,26	0,26
P1=9;P2=1	P1	0,91	0,09	0,96	0,04	0,89	0,05	0,05	0,90	0,05	0,05	0,89	0,04	0,04	0,04	0,86	0,05	0,05	0,05
	P2	0,15	0,85	0,49	0,51	0,11	0,01	0,88	0,27	0,37	0,37	0,11	0,01	0,01	0,88	0,20	0,27	0,27	0,27

Figure

Xen46a (evidence) true content	Euroformix 2PM DNA amount		NITZq 2PM DNA amount		Euroformix 3PM DNA amount			NITZq 3PM DNA amount			Euroformix 4PM DNA amount				NITZq 4PM DNA amount			
(real contributors and ratios)	POI	U	POI	U	POI	U	U	POI	U	U	POI	U	U	U	POI	U	U	U
P1=1;P3=1;P2=1	0,36 0,03	0,64 0,03	0,46 0,02	0,54 0,02	0,16 0,04	0,42 0,02	0,42 0,02	0,30 0,01	0,35 0,01	0,35 0,01	0,08 0,04	0,31 0,02	0,31 0,02	0,31 0,02	0,22 0,01	0,26 0,00	0,26 0,00	0,26 0,00
P1=5;P3=5;P2=1	0,34 0,04	0,66 0,04	0,42 0,01	0,58 0,01	0,14 0,03	0,43 0,02	0,43 0,02	0,27 0,01	0,37 0,01	0,37 0,01	0,06 0,04	0,31 0,01	0,31 0,01	0,31 0,01	0,20 0,01	0,27 0,00	0,27 0,00	0,27 0,00
P1=1;P3=1;P2=5	0,32 0,04	0,68 0,04	0,40 0,02	0,60 0,02	0,14 0,03	0,42 0,08	0,43 0,08	0,25 0,01	0,38 0,01	0,38 0,01	0,06 0,09	0,32 0,08	0,30 0,06	0,32 0,08	0,18 0,01	0,27 0,00	0,27 0,00	0,27 0,00

AVG(*)

std (**)

AVG(*)

std (**)

AVG(*)

std (**)

(*) Average amount calculated from 40 measurements obtained by placing as many unrelated individuals in the role of a POI

(**) its standard deviation

Figure

Xen13a (evidence) true content	Euroformix 2PM DNA amount		NITZq 2PM DNA amount		Euroformix 3PM DNA amount			NITZq 3PM DNA amount			Euroformix 4PM DNA amount				NITZq 4PM DNA amount			
(real contributors and ratios)	POI	U	POI	U	POI	U	U	POI	U	U	POI	U	U	U	POI	U	U	U
P1=1;P2=1	0,22	0,78	0,42	0,58	0,02	0,49	0,49	0,31	0,35	0,35	0,00	0,04	0,48	0,48	0,23	0,26	0,26	0,26
	0,04	0,04	0,02	0,02	0,02	0,01	0,01	0,02	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,01	0,01	0,01
P1=9;P2=1	0,19	0,81	0,36	0,64	0,03	0,07	0,89	0,22	0,39	0,39	0,02	0,05	0,07	0,87	0,16	0,28	0,28	0,28
	0,05	0,05	0,01	0,01	0,01	0,01	0,00	0,01	0,01	0,01	0,01	0,00	0,13	0,13	0,01	0,00	0,00	0,00

AVG(*)

std (**)

AVG(*)

std (**)

(*) Average amount calculated from 40 measurements obtained by placing as many unrelated individuals in the role of a POI

(**) its standard deviation

Figure

Xen46a (evidence) true content	Euroformix semicont QUANT				NITZq 2PM semicont QUANT				Euroformix semicont QUANT				NITZq 3PM semicont QUANT				Euroformix semicont QUANT				NITZq 4PM semicont QUANT			
(real contributors and ratios)	P1U/UU				P1UU/UUU				P1UUU/UUUU				P1UUU/UUUU				P1UUU/UUUU				P1UUU/UUUU			
P1=1;P3=1;P2=1	1,2	13,1	1,3	10,5	1,8	13,7	1,9	12,1	1,7	13,8	1,7	11,8	1,7	13,8	1,7	11,8	1,7	13,8	1,7	11,8	1,7	11,8	1,7	11,8
P1=5;P3=5;P2=1	1,2	18,6	1,3	11,6	1,8	19,1	1,9	14,4	1,7	19,5	1,7	14,3	1,7	19,5	1,7	14,3	1,7	19,5	1,7	14,3	1,7	14,3	1,7	14,3
P1=5;P3=1;P2=5	1,2	28,1	1,3	21,9	1,8	28,8	1,9	23,6	1,7	29,5	1,7	20,9	1,7	29,5	1,7	20,9	1,7	29,5	1,7	20,9	1,7	20,9	1,7	20,9
P1=1;P3=5;P2=5	1,2	1,7	1,3	-0,6	1,8	1,7	1,9	0,04	1,7	1,6	1,7	0,5	1,7	1,6	1,7	0,5	1,7	1,6	1,7	0,5	1,7	0,5	1,7	0,5
P1=1;P3=1;P2=5	1,2	44,6	1,3	36,4	1,8	45,3	1,9	37,2	1,7	46,4	1,7	33,0	1,7	46,4	1,7	33,0	1,7	46,4	1,7	33,0	1,7	33,0	1,7	33,0
P1=5;P3=1;P2=1	1,2	4,3	1,3	2,5	1,8	4,9	1,9	4,3	1,7	4,9	1,7	4,7	1,7	4,9	1,7	4,7	1,7	4,9	1,7	4,7	1,7	4,7	1,7	4,7
P1=1;P3=5;P2=1	1,2	1,8	1,3	-0,1	1,8	2,6	1,9	1,4	1,7	2,8	1,7	0,7	1,7	2,8	1,7	0,7	1,7	2,8	1,7	0,7	1,7	0,7	1,7	0,7
(real contributors and ratios)	P3U/UU				P3UU/UUU				P3UUU/UUUU				P3UUU/UUUU				P3UUU/UUUU				P3UUU/UUUU			
P1=1;P3=1;P2=1	0,9	10,8	1,1	8,4	1,7	11,1	1,8	10,2	1,6	11,1	1,7	9,7	1,6	11,1	1,7	9,7	1,6	11,1	1,7	9,7	1,6	9,7	1,6	9,7
P1=5;P3=5;P2=1	0,9	14,4	1,1	8,0	1,7	14,7	1,8	11,7	1,6	14,9	1,7	10,7	1,6	14,9	1,7	10,7	1,6	14,9	1,7	10,7	1,6	10,7	1,6	10,7
P1=5;P3=1;P2=5	0,9	2,2	1,1	-0,9	1,7	1,4	1,8	0,4	1,6	1,2	1,7	0,6	1,6	1,2	1,7	0,6	1,6	1,2	1,7	0,6	1,6	0,6	1,6	0,6
P1=1;P3=5;P2=5	0,9	23,0	1,1	18,5	1,7	23,5	1,8	19,6	1,6	24,0	1,7	17,5	1,6	24,0	1,7	17,5	1,6	24,0	1,7	17,5	1,6	17,5	1,6	17,5
P1=1;P3=1;P2=5	0,9	3,6	1,1	-0,6	1,7	2,0	1,8	2,0	1,6	2,3	1,7	0,9	1,6	2,3	1,7	0,9	1,6	2,3	1,7	0,9	1,6	0,9	1,6	0,9
P1=5;P3=1;P2=1	0,9	1,2	1,1	1,7	1,7	3,4	1,8	3,0	1,6	3,3	1,7	3,1	1,6	3,3	1,7	3,1	1,6	3,3	1,7	3,1	1,6	3,1	1,6	3,1
P1=1;P3=5;P2=1	0,9	39,2	1,1	31,6	1,7	39,7	1,8	32,1	1,6	40,8	1,7	28,7	1,6	40,8	1,7	28,7	1,6	40,8	1,7	28,7	1,6	28,7	1,6	28,7
(real contributors and ratios)	P2U/UU				P2UU/UUU				P2UUU/UUUU				P2UUU/UUUU				P2UUU/UUUU				P2UUU/UUUU			
P1=1;P3=1;P2=1	2,4	2,8	2,5	1,9	2,2	2,7	2,2	0,3	1,8	2,5	1,9	2,1	1,8	2,5	1,9	2,1	1,8	2,5	1,9	2,1	1,8	2,1	1,8	2,1
P1=5;P3=5;P2=1	2,4	2,1	2,5	-2,0	2,2	0,8	2,2	-2,6	1,8	0,2	1,9	-1,9	1,8	0,2	1,9	-1,9	1,8	0,2	1,9	-1,9	1,8	-1,9	1,8	-1,9
P1=5;P3=1;P2=5	2,4	4,5	2,5	2,4	2,2	4,9	2,2	2,1	1,8	4,8	1,9	3,7	1,8	4,8	1,9	3,7	1,8	4,8	1,9	3,7	1,8	3,7	1,8	3,7
P1=1;P3=5;P2=5	2,4	4,3	2,5	2,1	2,2	4,7	2,2	1,8	1,8	4,6	1,9	3,1	1,8	4,6	1,9	3,1	1,8	4,6	1,9	3,1	1,8	3,1	1,8	3,1
P1=1;P3=1;P2=5	2,4	1,4	2,5	-1,4	2,2	1,1	2,2	-1,7	1,8	1,2	1,9	-1,1	1,8	1,2	1,9	-1,1	1,8	1,2	1,9	-1,1	1,8	-1,1	1,8	-1,1
P1=5;P3=1;P2=1	2,4	18,3	2,5	15,6	2,2	18,3	2,2	15,9	1,8	18,3	1,9	14,4	1,8	18,3	1,9	14,4	1,8	18,3	1,9	14,4	1,8	14,4	1,8	14,4
P1=1;P3=5;P2=1	2,4	1,4	2,5	-1,4	2,2	0,4	2,2	-2,0	1,8	0,4	1,9	-1,9	1,8	0,4	1,9	-1,9	1,8	0,4	1,9	-1,9	1,8	-1,9	1,8	-1,9

Figure

Xen13a (evidence) true content (real contributors and ratios)	Euroformix		NITZq 2PM		Euroformix		NITZq 3PM		Euroformix		NITZq 4PM	
	semicont QUANT		semicont QUANT		semicont QUANT		semicont QUANT		semicont QUANT		semicont QUANT	
	P1U/UU				P1UU/UUU				P1UUU/UUUU			
P1=1;P2=1	7,11	27,67	9,1	23,5	6,77	28,05	8,6	23,8	6,35	28,03	8,2	23,1
P1=1;P2=1	7,21	12,68	9,1	7,4	6,90	14,95	8,6	9,8	6,49	14,95	8,2	10,2
P1=1;P2=1	6,63	3,79	9,1	3,0	9,70	17,75	8,6	4,7	6,01	17,82	8,4	31,3
P1=1;P2=1	7,30	57,88	8,8	40,7	7,02	57,86	8,3	41,4	6,63	57,86	7,8	37,5
P1=1;P2=1	7,11	58,68	8,8	48,8	6,77	58,66	8,3	47,0	6,35	58,67	7,8	42,1
(real contributors and ratios)	P2U/UU				P2UU/UUU				P2UUU/UUUU			
P1=1;P2=1	7,31	10,26	10,4	6,3	7,15	10,72	9,0	5,5	6,50	10,73	8,4	7,6
P1=1;P2=1	8,40	28,09	10,4	18,5	7,27	26,82	9,0	19,3	6,63	26,82	8,4	19,0
P1=1;P2=1	7,83	52,02	10,4	34,9	6,77	52,18	9,0	35,1	6,15	52,16	8,4	31,3
P1=1;P2=1	8,50	5,82	9,9	2,5	7,39	20,42	8,6	2,9	6,77	20,47	7,9	2,5
P1=1;P2=1	8,31	4,57	10,0	1,6	7,15	11,23	8,6	2,7	6,50	10,34	7,9	1,5

Figure

Xen46a (evidence)		POI U/UU						POI UU/UUU						POI UUU/UUUU					
true content				percentiles						percentiles						percentiles			
P1=1;P3=1;P2=1		min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50
EUROFORMIX	semicont	-10,38	-1,23	-1,49	-1,92	-3,02	-4,85	-6,81	-0,86	-0,98	-1,20	-1,53	-3,15	-5,80	-0,23	-0,58	-1,17	-1,32	-2,62
	quant	-12,68	-1,92	-2,79	-4,15	-4,35	-7,65	-4,75	0,97	0,36	-1,28	-1,67	-2,91	-1,85	1,65	1,12	-0,36	-0,54	-1,36
NITZ	semicont	-21,67	-11,93	-11,95	-12,56	-12,82	-17,19	-17,35	-10,84	-10,88	-11,06	-11,16	-14,81	-18,18	-12,01	-12,03	-12,08	-12,10	-15,97
	quant	-31,43	-18,13	-18,17	-18,24	-19,41	-25,80	-36,55	-26,72	-26,91	-27,45	-28,03	-31,81	-37,34	-24,82	-24,96	-25,96	-27,09	-31,76
P1=5;P3=5;P2=1		min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50
EUROFORMIX	semicont	-10,38	-1,23	-1,49	-1,92	-3,02	-4,85	-6,81	-0,86	-0,98	-1,20	-1,53	-3,15	-5,80	-0,17	-0,19	-1,09	-1,32	-2,62
	quant	-13,07	-1,27	-1,31	-2,64	-4,79	-6,90	-4,69	2,03	1,12	-0,84	-1,15	-2,37	-1,32	2,76	1,74	0,16	-0,28	-0,85
NITZ	semicont	-21,76	-11,65	-11,88	-12,65	-12,81	-17,06	-16,91	-10,49	-10,57	-10,79	-10,88	-14,34	-17,49	-11,55	-11,58	-11,65	-11,70	-15,28
	quant	-30,92	-18,39	-18,55	-19,20	-19,67	-25,71	-34,76	-25,74	-25,79	-26,05	-26,82	-30,26	-38,19	-26,82	-26,84	-27,04	-27,53	-32,77
P1=1;P3=1;P2=5		min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50	min	max	0,99	0,95	0,90	0,50
EUROFORMIX	semicont	-10,38	-1,23	-1,49	-1,92	-3,02	-4,85	-6,81	-0,86	-0,98	-1,20	-1,53	-3,15	-5,80	-0,23	-0,58	-1,17	-1,32	-2,78
	quant	-12,30	-0,93	-0,96	-2,67	-3,86	-6,43	-3,98	2,37	1,32	-0,66	-0,86	-2,17	-1,34	2,85	1,94	0,16	-0,09	-0,74
NITZ	semicont	-22,11	-12,02	-12,13	-12,77	-13,15	-17,59	-17,35	-10,88	-10,90	-11,09	-11,24	-14,76	-17,97	-11,94	-11,96	-12,02	-12,06	-15,81
	quant	-26,72	-12,68	-12,82	-13,07	-14,93	-20,97	-34,54	-24,44	-24,54	-25,79	-25,96	-29,72	-37,85	-25,08	-25,20	-25,73	-26,92	-31,42

Figure

Xen13a (evidence)		POI U/UU					
true content				percentiles			
P1=1;P2=1		min	max	0,99	0,95	0,90	0,50
EUROFORMIX	semicont	-20,27	-5,45	-5,63	-6,14	-7,03	-11,20
	quant	-14,85	-4,80	-5,07	-5,83	-6,55	-9,64
NITZ	semicont	-84,64	-56,62	-58,40	-63,24	-65,54	-76,18
	quant	-85,39	-50,75	-53,24	-60,24	-62,56	-75,77
P1=9;P2=1		min	max	0,99	0,95	0,90	0,50
EUROFORMIX	semicont	-21,27	-5,45	-5,63	-6,14	-7,03	-11,20
	quant	-24,80	-7,24	-8,42	-10,88	-14,29	-17,98
NITZ	semicont	-90,25	-57,86	-59,71	-65,26	-67,43	-80,12
	quant	-86,39	-47,15	-49,89	-58,28	-60,96	-74,07

POI UU/UUU							
				percentiles			
		min	max	0,99	0,95	0,90	0,50
		-18,52	-2,88	-3,32	-4,56	-5,90	-9,22
		-0,82	0,54	0,54	0,51	0,21	-0,45
		-53,36	-19,77	-19,81	-23,20	-26,59	-34,31
		-56,90	-18,78	-19,34	-25,85	-27,83	-37,16
		min	max	0,99	0,95	0,90	0,50
		-18,52	-2,88	-3,32	-4,56	-5,47	-9,22
		-2,59	0,83	0,61	0,22	-0,52	-1,17
		-76,13	-54,26	-56,25	-61,34	-63,38	-70,25
		-86,52	-65,23	-66,48	-71,19	-72,37	-79,50

POI UUU/UUUU							
				percentiles			
		min	max	0,99	0,95	0,90	0,50
		-17,88	-2,26	-2,74	-4,41	-5,66	-8,78
		-0,02	0,53	0,53	0,49	0,26	-0,01
		-53,60	-19,24	-19,49	-22,79	-26,28	-33,78
		-56,61	-18,76	-19,34	-25,68	-26,85	-36,98
		min	max	0,99	0,95	0,90	0,50
		-17,88	-2,26	-2,74	-4,41	-5,66	-8,78
		-0,95	0,68	0,58	0,34	0,01	-0,40
		-76,68	-54,99	-57,08	-62,02	-63,87	-70,75
		-89,79	-67,10	-67,82	-73,34	-74,50	-82,17

Figure

Figure

Figure

Figure

Figure