Quality of *NAT2* Genotyping with Restriction Fragment Length Polymorphism Using DNA Isolated from Frozen Urine[1]

**Fränzel J. B. van Duijnhoven, Olga L. van der Hel,
Rob B. van der Luijt, H. Bas Bueno de Mesquita,
Paul A. H. van Noord, and Petra H. M. Peeters[2]**

Julius Center for General Practice and Patient Oriented Research
[F. J. B. v. D., O. L. v. d. H., P. A. H. v. N., P. H. M. P.] and Department of
Medical Genetics [F. J. B. v. D., R. B. v. d. L.], University Medical Center,
3508 GA Utrecht, the Netherlands, and National Institute of Public Health and
Environmental Protection, 3720 BA Bilthoven, the Netherlands
[H. B. B. d. M.]

Abstract

**In large studies and under field conditions common to
epidemiological research, factors outside of and inside
the laboratory can introduce misclassification of genetic
susceptibility markers. Few reports have been made on
the accuracy of genotyping individuals using DNA
extracted from frozen urine that was stored for ~20
years. This study was performed to determine the
reproducibility and accuracy of *N-acetyltransferase 2*
(*NAT2*) genotyping by RFLP analysis using DNA from
stored urine. To obtain long-term frozen urine and blood
samples from the same person, the databases of two large
prospective studies were linked by name and date of
birth. Six polymorphisms within the coding region of
NAT2 were determined in 65 urine and blood samples
after which, genotypes and imputed phenotypes (rapid,
slow) were derived. To test reproducibility, all of the six
polymorphisms were determined twice in 47 urine-blood
pairs. Reproducibility of imputed phenotypes was 91.5%
in urine samples and 97.9% in blood samples. To test
accuracy, results for all six polymorphisms were
compared between urine and blood DNA. All of the κ's
were at least 0.85 except one. Identical results for all six
polymorphisms were seen in 78.5% of urine-blood pairs.
Taking blood samples as a reference standard, rapid
acetylators were classified as rapid in 97% of subjects
(95% confidence interval, 90–100%), and slow acetylators
were classified as slow also in 97% of subjects (95%
confidence interval, 91–100%), when using urine. This
study shows that stored urine samples can be used for
DNA genotyping in large cohort studies, when blood
samples are not available.**

Introduction

In the past, human tissues such as urine samples and nails (1, 2)
have been stored for prospective epidemiological studies. The
initial goal was to use these biological specimens for protein
analysis or the determination of hormone levels to study their
relationships to the occurrence of chronic diseases in the future.
Genotyping of individuals is usually done with DNA isolated
from blood leukocytes, but blood samples have not always been
collected in large cohort studies. Because it became clear that
DNA could be extracted from human material other than blood
such as urine, the possibility of using these specimens for
genetic epidemiological analyses in population-based studies
arose. Although the amount of DNA extracted from these
specimens may be low, it is sufficient for PCR (3, 4), which
makes it possible to perform genetic analyses.

In field conditions common to epidemiological research,
several factors can induce misclassification of genetic markers,
such as a mix-up in sample collection, errors in processing, and
laboratory factors, *e.g.,* the failure to recognize a variant that
contributes to the genotype of interest, false priming in PCR-
based arrays, and the erroneous recognition of a pseudogene.
Finally, coding errors at any point during sample collection,
laboratory analysis, or data processing may occur (5).

Urine samples have been used earlier as a source for DNA
(4, 6). In a large prospective study (DOM[3]; Ref. 7), we col-
lected urine samples (1974–1986) from thousands of women
and stored them at −20°C for future use. During follow-up,
some of these women were diagnosed with cancer. In a nested
case-control study, we plan to isolate DNA from relevant study
subjects. One of the interests is to study polymorphic low-risk
genes in interaction with environmental factors and the occur-
rence of cancer. For this purpose, genes coding for enzymes
involved in the metabolism of carcinogens, such as *N*-acetyl-
transferases, are of interest. These enzymes are genetically
polymorphic. In the present study, we assessed the reproduc-
ibility and accuracy of *NAT2* genotyping, with the RFLP
method in DNA obtained from frozen urine samples (stored for
~20 years), in 65 subjects from whom blood samples had also
been collected.

Materials and Methods

Study Population. Urine samples were collected from
women, who participated in the DOM study, a study to assess
breast cancer screening efficacy. Women (n ≈ 30,000) who
were born between 1911 and 1945 and were living in Utrecht
and surroundings (the Netherlands), were enrolled between
1974 and 1986 for the DOM study. Apart from a screening

Table 1 Characteristics of PCRs for fragment analysis and *NAT2* analysis

Locus	Fragment length (bp)	Program	Number of cycles	Primer sequence	Primer orientation
D3S1263	231–249	10 min 95°C, (30 s 95°C; 30 s 55°C; 30 s 72°C), 30 min 72°C	33	5′-ctgttgacccattgataccc	forward
				5′-taaaatcacagcaggggttc	reverse
D20S119	101–119			5′-ctgacacagtttcagtatctctatc	forward
				5′-tttccagatttaggggtgtatg	reverse
NAT2: 230 standard	350	10 min 95°C, (30 s 95°C; 30 s 56°C; 30 s 72°C), 10 min 72°C	20	5′-gtcacacgaggaaatcaaatgc	forward
				5′-acccagcatcgacaatgtaattcctgccctca[a]	reverse
NAT2: 230 nested	230	10 min 95°C, (30 s 95°C; 30 s 56°C; 30 s 72°C), 10 min 72°C	34	5′-ggagttgggcttagaggc	forward
				5′-acccagcatcgacaatgtaattcctgccctca[a]	reverse
NAT2: 524 standard	547	10 min 95°C, (1 min 95°C; 1 min 60°C; 2 min 72°C), 10 min 72°C	20	5′-gctgggtctggaagctcctc	forward
				5′-ttgggtgatacatacacaggg	reverse
NAT2: 524 nested	524	10 min 95°C, (30 s 95°C; 30 s 54°C; 1 min 72°C), 10 min 72°C	34	5′-ccagatgtggcagcctcta	forward
				5′-ggtgatacatacacaagggttt	reverse

[a] Mismatch primer.

mammography, lifestyle questionnaires were filled out and overnight urine samples were collected. Approximately 200 ml of urine was stored for each participant at 4°C for 8 h and was then transported to a −20°C freezer. The samples were kept at −20°C until DNA isolation was performed, which was ∼20 years later (range, 11–23 years).

Some of the DOM participants were also included in the Prospect-EPIC study, which is a Dutch cohort participating in the EPIC study (8). This study included ∼17,000 women who were born between 1923 and 1947 and were living in Utrecht and surroundings (the Netherlands). Between 1993 and 1997, questionnaires were filled out and blood samples were collected, which were stored at −196°C. After fractionating, some (whole) blood remained and was stored at −20°C for every participant after 1995.

To identify persons with urine and blood samples, data of the DOM cohort and the Prospect-EPIC cohort were linked. Linkage was done on surname with initials and date of birth. This resulted in the identification of 67 persons who had donated both a urine and a blood sample.

DNA Isolation from Frozen Urine. The urine samples were thawed overnight at room temperature and mixed vigorously, before 50 ml were removed. Cells and nuclei were centrifuged at $2,000 \times g$ for 15 min and then washed twice in PBS at $2,000 \times g$ for 15 min. After resuspension in 500 μl of PBS, cells and nuclei were spun down for 5 min. at $12,000 \times g$. The pellet was resuspended in 300 μl of cell lysis buffer and was incubated overnight at 55°C. Proteins were precipitated by the addition of 100 μl of protein precipitation solution (6 M NaCl) and the precipitate was pelleted by a 5-min centrifugation at $12,000 \times g$. The supernatant was treated with one volume absolute ethanol, and DNA was centrifuged at $12,000 \times g$ for 5 min. The pellet was washed with 70% ethanol and centrifuged for 5 min at $12,000 \times g$. After drying, the DNA pellet was resuspended in 40 μl of TE [10 mM Tris, 1 mM EDTA (pH 7.6)]. Previously, we studied quality and quantity of DNA isolated from comparable urine samples (9). Because the range of DNA concentration was not normally distributed

(3 of 48 samples were outliers), the 25th percentile was 4 ng/μl, the 50th percentile (median) was 14 ng/μl, and the 75th percentile was 28 ng/μl. This previous study showed that although DNA amount was variable, 26–89% of the samples, depending both on the length of the PCR amplicon and on PCR conditions, yielded a visible PCR product. Furthermore, the percentage of visible PCR product on agarose gel was lower for low amounts of DNA (0–0.1 μl) than for higher amounts of DNA (0.1–20 μg; Ref. 9).

DNA Isolation from Frozen Blood. The blood samples were thawed overnight at 4°C and mixed vigorously. The DNA isolation was performed with a QIAamp DNA Mini kit (Qiagen, Leusden, the Netherlands) according to the manufacturer's Blood and Body Fluid Spin protocol. The quantity of DNA extracted from blood was measured using a UV spectrophotometer. The concentration ranged from 12 ng/μl to 55 ng/μl. The mean was 27 ng/μl.

Fragment Analysis. Because administrative twins exist, two polymorphic markers, *D3S1263* and *D20S119*, were used for fragment analysis to determine whether the urine and blood samples were indeed from the same person. A 10-μl multiplex PCR was performed, which contained Perkin-Elmer Gold buffer [50 mM KCl, 15 mM Tris-HCl (pH 8.0)], 2.5 mM MgCl$_2$, 0.2 mM each nucleotide (Pharmacia, Biotech, Uppsala, Sweden), 25 ng of each primer (Isogen Bioscience BV, Maarssen, the Netherlands), 0.4 units of Amplitaq DNA polymerase (Perkin-Elmer), and 2 μl of DNA from urine or blood. PCR conditions are given in Table 1.

After PCR, 4 μl of loading mix [43:1 (v/v) Hi-Di Formamide:GeneScan 500 XL Rox Size standard] was added to 1 μl of undiluted PCR product. Fragment analysis was performed on the ABI Prism 3700 DNA Analyzer (Perkin-Elmer) according to the manufacturer's instructions. Analysis was done with GeneScan analysis 3.1 and Genotyper 2.1 software. The results on the markers were not identical for two urine-blood pairs. These two were removed from the analyses (Fig. 1).
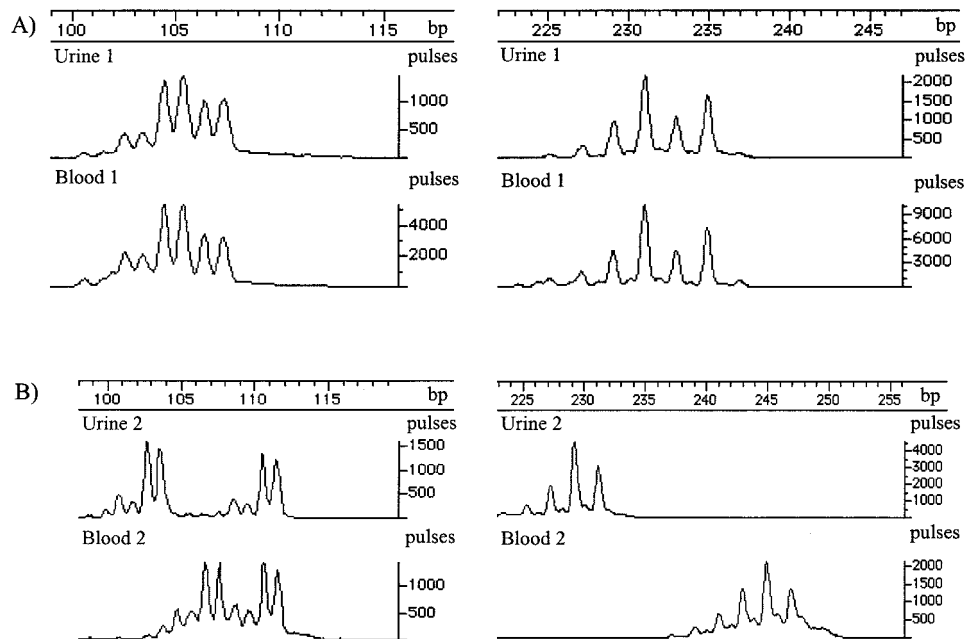
Fig. 1. Markerpattern for fragment analysis with *D20S119* and *D3S1263*. *Left*: *D20S119*. *Right*: *D3S1263*. *A*, identical pattern for linked urine and blood sample 1. *B*, not identical pattern for linked urine and blood sample 2. Fragment lengths are indicated in bp. Intensity of markerpatterns is indicated in number of pulses.

NAT2 **Analysis.** To test the quality of *NAT2* genotyping in urine, genotyping was performed on DNA that was isolated from both urine and blood samples from the same individual. A PCR-RFLP method was used to analyze six polymorphisms (*C282T, T341C, C481T, G590A, A803G,* and *G857A*) within the coding region of *NAT2*.

First, a standard PCR was performed. To yield enough PCR product for RFLP, the product of the first PCR was used as a template for a nested PCR. All of the PCR products of the standard PCRs were stored at 4°C and used later in another experiment to test the reproducibility. Two different sets of PCR primers were used to obtain *NAT2* fragments of 230 and 524 bp for further analysis. All of the PCRs were performed in 25-μl volumes, which contained Perkin-Elmer buffer II [50 mM KCl, 10 mM Tris-HCl (pH 8.3)], 2.5 mM MgCl$_2$, 0.2 mM each nucleotide (Pharmacia, Biotech), 0.2 mg/ml BSA, 10 pmol of each primer (Isogen Bioscience BV, Maarssen, the Netherlands), and 0.05 units of Amplitaq Gold polymerase (Perkin-Elmer). For PCRs, 2 μl of DNA from urine, 1 μl of DNA from blood, or 1 μl of PCR product was used. To generate a *Dde*I restriction site for the polymorphism at nucleotide position 341, a mismatch primer was used. Primer sequences and PCR conditions are shown in Table 1.

The 230-bp fragment contains two *NAT2* polymorphisms (*C282T* and *T341C*) and is digested with restriction enzymes *Fok*I (0.92 units; New England Biolabs, Leusden, the Netherlands) and *Dde*I (2.3 units; Boehringer Mannheim, Mannheim, Germany). The four other *NAT2* polymorphisms within the 524-bp fragment (*C481T, G590A, A803G,* and *G857A*) are determined with the use of restriction enzymes *Taq*I (4.6 U; New England Biolabs, Leusden, the Netherlands), *Kpn*I, *Dde*I, and *Bam*HI (2.3 units; Boehringer Mannheim). All of the digestion reactions were performed in 10 μl with the most optimal buffer for every restriction enzyme according to the manufacturer. Restriction products were electrophoresed in a 3% agarose gel and restriction fragment patterns were visualized by ethidium bromide staining. Expected basepair sizes are shown in Table 2. The samples were classified for each polymorphism

Table 2 Expected bp sizes for the determination of six polymorphisms in the *NAT2* gene with RFLP

Fragment length (bp)	Nucleotide substitution	Restriction enzyme	Expected bp sizes
230	C282T	*Fok*I	230, 125, 105
	T341T	*Dde*I	220, 188, 32, 10
524	C481T	*Taq*I	371, 201, 170, 153
	G590A	*Kpn*I	524, 430, 94
	A803G	*Dde*I	345, 135, 112, 44, 23
	G857A	*Bam*HI	524, 470, 54

as homozygous wild type, heterozygous, or homozygous mutant by two observers independently (F.J.B.v.D. and O.L.v.d.H.). The concordance rate was 97%. The few differences in interpretation that occurred could always be resolved by discussion. To check this assay, five urine and blood pairs were sequenced. The sequence data of these samples were completely concordant with the RFLP data.

Imputed Phenotype Classification. Depending on the six polymorphisms, both *NAT2* alleles were determined for urine and blood samples separately. *NAT2* genotypes were derived from these alleles. Because no probe drug was used, it was not possible to determine actual phenotype. Urine was collected 20 years ago (range, 11–23 years) and it was not feasible to contact the women a second time. Therefore, genotypes were used to assess imputed phenotypes. The samples were classified as rapid acetylator phenotype when the *NAT2* genotype consisted of at least one of the following haplotypes: *NAT2*4, NAT2*12A, NAT2*12B, NAT2*12C,* or *NAT2*13*. All other *NAT2* genotypes were classified as slow acetylator phenotype.

Statistical Analyses. To assess reproducibility, PCRs for all six of the polymorphisms were done twice in 47 urine and blood samples. Observers were blinded for the initial allele determination. Unfortunately, 18 of 65 urine-blood pairs were excluded from the reproducibility tests because of an error that

Table 3   Agreement of RFLP results per polymorphism between urine and blood samples and comparison of allele frequencies based on DNA from urine and blood samples with those observed for a Caucasian population[a]

Nucleotide position	Identical (%)	κ	Allele frequencies in urine (%)	Allele frequencies in blood (%)	Allele frequencies in literature (%)
282	89.2	0.85	30.0	29.2	30.8
341	93.8	0.90	42.2	41.5	46.5
481	89.2	0.85	39.2	36.9	42.5
590	92.3	0.88	24.6	26.9	27.8
803	95.4	0.94	40.8	41.5	42.4
857	98.5	0.66	0.8	1.5	1.3
Average	93.1				

[a] Ref. 12.

occurred during storage of the samples after the accuracy tests. Percentages of agreement between first and second RFLP results were computed.

To test accuracy, all of the six polymorphisms were determined in 65 samples of urine and blood. When determining the polymorphisms in blood samples, observers were blinded for initial polymorphism determination in urine samples. For each polymorphism, a weighted κ was computed (10). Next, genotype and imputed phenotype were determined for each person based on DNA obtained from urine and subsequently based on DNA obtained from blood. Taking blood as the reference standard, the sensitivity rate (those correctly classified as a rapid acetylator) and specificity rate (those correctly classified as a slow acetylator) were computed for imputed phenotypes based on DNA extracted from urine. Next, we calculated the bias in the estimation of the RR caused by a misclassification of the imputed phenotype. For this, we assumed a (nested) case-control study, in which the classification of rapid or slow acetylator type was based on DNA obtained from urine samples. We may then assume that the misclassification is nondifferential, if the observers were blinded as to case or control status. The following formula (11) was used to calculate the ORR as a function of the TRR, the true prevalence of the risk factor (*Prev*), and the sensitivity (*Se*) and specificity (*Sp*) of the assay.

$$ORR = \frac{\begin{array}{c}[Se \times \text{TRR} \times Prev + (1 - Sp)(1 - Prev)] \\ \times [(1 - Se)Prev + Sp(1 - Prev)]\end{array}}{\begin{array}{c}[Se \times Prev + (1 - Sp)(1 - Prev)] \times [(1 \\ - Se\text{TRR} \times Prev + Sp(1 - Prev)]\end{array}}$$

We calculated ORRs for several TRRs (constant prevalence = 0.4) and for several prevalences (constant TRR = 2).

Results

Reproducibility. By performing nested PCRs and RFLP again and determining all of the six polymorphisms of *NAT2* for a second time, the reproducibility of *NAT2* genotyping with RFLP in 47 urine samples as well as 47 blood samples was tested.

Differences between the first and second time of *NAT2* genotyping were seen in 8 (2.8%) of 282 polymorphism determinations (47 samples × 6 polymorphisms per sample). This resulted in a different genotype in 7 (14.9%) of 47 urine samples (6 samples differed in 1 polymorphism and 1 sample differed in 2 polymorphisms). Although different genotypes were seen in seven urine samples, three of these seven samples still showed the same imputed phenotype. Eventually, 4 (8.5%) of 47 urine samples showed a difference in imputed phenotype.

Table 4   Comparison of imputed *NAT2* phenotype classification between urine and blood samples

Samples were classified as rapid acetylator, slow acetylator, or unknown.

Urine	Blood			
	Rapid	Slow	Unknown	Total
Rapid	29	1	1	31
Slow	1	30	0	31
Unknown	2	1	0	3
Total	32	32	1	65

One (0.4%) of 282 polymorphisms in blood samples showed a difference, resulting in a different genotype and an imputed phenotype in 1 (2.1%) of 47 blood samples.

Accuracy. Six *NAT2* polymorphisms were determined for all 65 urine and blood samples. The percentages of agreement per polymorphism between urine and blood samples showed an average of 93%. κ for each polymorphism was >85% for all but one (*i.e.,* 66%). Allele frequencies per polymorphism for DNA obtained from urine and blood samples were comparable with those observed for a Caucasian population (Ref. 12; Table 3).

Fifty-one (78.5%) of 65 urine and blood pairs were identical for all six polymorphisms. Fourteen pairs were not identical, ranging from differences for 1 to 4 polymorphisms. One urine-blood pair differed in four polymorphisms, three pairs differed in three polymorphisms, four pairs differed in two polymorphisms, and six pairs differed in one polymorphism.

After determining the *NAT2* polymorphisms, *NAT2* genotypes and imputed phenotypes were calculated for most of the samples. It was not possible to assign genotypes and, thus, imputed phenotypes in three urine samples and one blood sample, because the combination of the six polymorphisms for these samples was not described in literature.[4] NAT2 phenotypes for urine samples and blood samples were compared with results obtained from blood samples and are shown in Table 4. A difference in imputed phenotype occurs in 6 (9.2%) of 65 samples. When unknown samples are excluded, differences occur in 2 (3.3%) of 61 samples.

Taking phenotype classification based on DNA obtained from blood as the reference standard, sensitivity rate (those with rapid acetylator type classified as rapid: 29 of 30) and specificity rate (those with slow acetylator type classified as slow: 30 of 31) were both 0.97, when using DNA from urine (Table 4). Using these sensitivity and specificity rates, Fig. 2

---

[4] Internet address: www.louisville.edu/medschool/pharmacology/NAT.html.
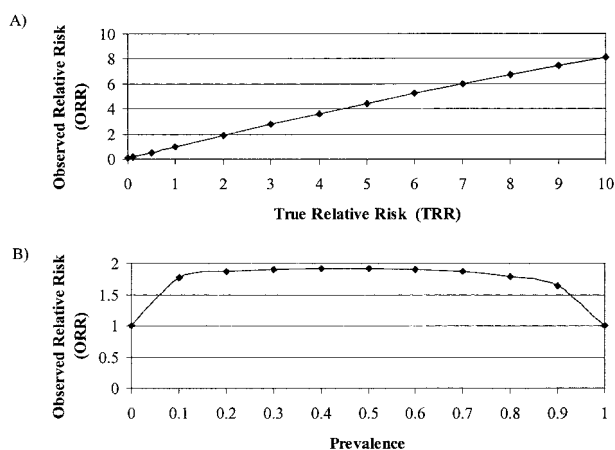
A)

B)

Fig. 2. ORR, given assay sensitivity = 0.97, assay specificity = 0.97, prevalence = 0.4, and TRR = 2, unless otherwise indicated. *A*, ORR as a function of TRR. *B*, ORR as a function of the prevalence.

shows the ORR for several values of the prevalence (of rapid acetylators) and TRRs. For example, with a prevalence of 0.4 of rapid acetylators in the population and a TRR of 2, we will observe a RR of 1.92 when we use DNA obtained from urine (Fig. 2).

Discussion

Comparison of the determination of six polymorphisms in the *NAT2* gene in DNA obtained from urine and blood showed differences in 14 of 65 urine-blood pairs. The κ for each of the six polymorphisms was >0.85 with the exception of one, which was attributable to a low frequency of the mutant allele. Taking blood samples as the reference standard, rapid acetylators were classified as rapid in 97% of subjects (sensitivity) and slow acetylators were classified as slow also in 97% of subjects (specificity), when urine samples were used.

We had the unique possibility to test accuracy of *NAT2* genotyping using DNA that was isolated from urine samples, which were frozen and stored for many years. In this way our study takes into account all of the conditions that are relevant to field work in epidemiological biomarker studies and that may introduce misclassification of the biomarker. These are: (*a*) possible errors in the identification code and storage, when large numbers are involved; (*b*) a loss of quality and quantity of DNA because of storage; (*c*) errors in the linkage of data sets; and (*d*) errors in identifying and retrieving both the urine and blood samples from the −20°C storage facility and errors during isolation of DNA and during the laboratory tests.

Although epidemiological studies make a great effort to minimize biomarker misclassification, the above mistakes will most likely occur in population-based studies. In this study, we have come across some errors, which may result in misclassification, despite our utmost carefulness. Although all of the urine-blood pairs were from the same person according to identifying data, polymorphic markers showed that 2 of 67 matched samples were not from the same individual. This can be caused by a linkage error or a mix-up in the storage of either urine or blood samples that were collected years before. Another explanation is that an error occurred in the process of extracting DNA from urine or blood or in the additional handling of this DNA in fragment analysis.

The reproducibility of imputed phenotype was not 100%

(91.5% in urine samples and 97.9% in blood samples), which could have several reasons. It is not possible to know whether it is a problem in the assay (faint bands, incomplete digestion, poor gel resolution) or a technician error (contamination or sample mix-up).

Another 14 urine-blood pairs showed differences in *NAT2* genotyping results, although they seemed to be from the same person according to administrative data and the identity of two polymorphic markers. However, 2 of ∼1000 persons will show the same identity when using these two markers, although they are different persons. It is unlikely that this occurred in 14 of 67 samples. However, if they were classified incorrectly as the same individual and we had excluded these individuals, results for reproducibility and accuracy would have been better. We think that these 14 urine-blood samples were, in fact, from the same person and that the differences were caused by incorrect classifying polymorphisms attributable to low amounts of DNA extracted from urine.

No genotype and, therefore, no imputed phenotype could be established for 3 of 65 urine samples and 1 of 65 blood samples. It is possible that an error occurred in the determination of one or more polymorphisms in these samples, which resulted in an unknown genotype and phenotype. Another explanation is that new genotypes have been found with these samples. However, this is unlikely because these samples showed differences in genotype between urine and blood DNA.

To compute the bias in the estimation of the RR attributable to misclassification of the rapid acetylator type, results of blood samples were taken as the reference standard. This is not a perfect reference standard, because reproducibility of genotyping of blood DNA was not 100 but 97.9% (1 of 47 samples showed a different acetylator type). For the classification of imputed phenotypes for rapid acetylators, six polymorphisms are involved and each one will be at risk for misclassification. When lesser loci are involved, the risk of misclassification of imputed phenotype, which is based on the results of combined loci, will be lower. If another gene with fewer polymorphisms is investigated, it might be expected that the occurrence of misclassification of imputed phenotypes is less.

Results were based on DNA obtained from female urine samples. It was reported before that DNA extraction from male urine samples will yield less DNA (6). Subsequently, for males, results might be different.

Smoking (yes/no) and older women (46 years of age and older) were slightly overrepresented in inaccurate samples ($n = 14$) when compared with accurate samples ($n = 51$). However, no firm conclusions can be drawn, because the numbers are very small.

Although the reproducibility of *NAT2* genotyping results with DNA from urine samples was not 100% and differences were observed when using DNA from blood samples, the sensitivity and specificity rates were high: 97% of subjects were classified correctly for acetylator type when urine samples were used instead of blood samples (95% confidence interval, 90–100% for sensitivity and 91–100% for specificity). Therefore, at least in women, stored urine samples can be used for DNA genotyping in large cohort studies, when blood samples are not available.

References

1. van Noord, P. A. H., Maas, M. J., van der Tweel, I., and Collette, C. Selenium and the risk of postmenopausal breast cancer in the DOM cohort. Breast Cancer Res. Treat., *25:* 11–19, 1993.

2. van den Brandt, P., Goldbohm, R. A., van 't Veer, P., Volovics, A., Hermus, R. J. J., and Sturmans, F. A large-scale prospective cohort study on diet and cancer in the Netherlands. J. Clin. Epidemiol., *43:* 285–295, 1990.

3. Lench, N., Stanier, P., and Williamson, R. Simple non-invasive method to obtain DNA for gene analysis. Lancet, *1:* 1356–1358, 1988.

4. Yokota, M., Tatsumi, N., Tsuda, I., Takubo, T., and Hiyoshi, M. DNA extraction from human urinary sediment. J. Clin. Lab. Anal., *12:* 88–91, 1998.

5. Rothman, N., Stewart, W. F., Caporaso, N., and Hayes, R. B. Misclassification of genetic susceptibility biomarkers: implications for case-control studies and cross-population comparisons. Cancer Epidemiol. Biomark. Prev., *2:* 299–303, 1993.

6. Prinz, M., Grellner, W., and Schmitt, C. DNA typing of urine samples following several years of storage. Int. J. Leg. Med., *106:* 75–79, 1993.

7. Waard, F., Rombach, J. J., Collette, H. J. A., and Honing, C. The DOM project for the early detection of breast cancer, Utrecht, the Netherlands. J. Chronic Dis., *34:* 1–44, 1984.

8. Keinan-Boker, L., van Noord, P. A. H., van der Schouw, Y. T., Koot, V. C. M., Bueno de Mesquita, H. B., Riboli, E., Grobbee, D. E., and Peeters, P. H. M. Prospect-EPIC Utrecht: study design and characteristics of the cohort population. Eur. J. Epidemiol., in press, 2002.

9. van der Hel, O. L., van der Luijt, R. B., Bueno de Mesquita, H. B., van Noord, P. A. H., Slothouber, B., Roest, M., Schouw, Y. T., Grobbee, D. E., Pearson, P. L., and Peeters, P. H. M. Quality and quantity of DNA isolated from frozen urine in population-based research. Anal. Biochem., *304:* 206–211, 2002.

10. Fleiss, J. L. Statistical methods for rates and proportions, pp. 143–147. New York: Wiley, 1993.

11. Flegal, K. M., Brownie, C., and Haas, J. D. The effects of exposure misclassification on estimates of relative risk. Am. J. Epidemiol., *123:* 736–751, 1986.

12. Cascorbi, I., Drakoulis, N., Brockmoller, J., Maurer, A., Sperling, K., and Roots, I. Arylamine *N*-acetyltransferase (*NAT2*) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. Am. J. Hum. Genet., *57:* 581–592, 1995.