Addressing reproductive risk in consanguineous couples

Teeuw, M.E.

2015

document version
Publisher's PDF, also known as Version of record

CHAPTER 7

Total pathogenic allele frequency of autosomal recessive MEFV mutations causing familial Mediterranean fever in Tunisia and Morocco

Teeuw ME/ Kelmemi W, Jonker MA, Kharrat M, Lariani I, Laarabi FZ, Hama I, Henneman L, Sefiani A, Chaabouni H, Ten Kate LP, Cornel MC.
Submitted

ABSTRACT

The total frequency of pathogenic alleles causing an autosomal recessive disorder can be estimated from mutational data on affected children of consanguineous parents and the inbreeding coefficient of the patients by means of a maximum likelihood method. In this chapter this method is applied to estimate the allele frequencies of bi-allelic autosomal recessive *MEFV* mutations in Tunisia and Morocco based on data of 293 Tunisian patients and 199 Moroccan patients with familial Mediterranean fever (FMF). The data were partly inconclusive, as in FMF patients molecular testing often fails to show two *MEFV* mutations. First, *q* was estimated based on individuals with bi-allelic mutations. Second, a fictive allele was modelled for all missing alleles and used for estimating *q*. The two estimates of total allele frequency for autosomal recessive *MEFV* FMF in Tunisia are 0.01040 (95% CI: 0.00381-0.02540) and 0.00949 (95% CI: 0.00460-0.01831). For the Moroccan data these estimates are (by lack of bi-allelic heterozygous patients) 0.00000 (95% CI: 0.00000-0.01326) and 0.002866 (95% CI:0.00042-0.0118). Our results show that by using a maximum likelihood method, the total allele frequency of an autosomal recessive disorder can be estimated – provided that reliable and sufficient data are available – even if some mutations are not identified.

INTRODUCTION

Information on the frequency of an autosomal recessive disorder in a particular population can be important for various reasons. In clinical practice, it can be helpful in diagnostic reasoning where the disease frequency codetermines its likelihood. After deducing total pathogenic allele and carrier frequency, it can also aid risk calculations in genetic counselling, for instance in couples who wish to know the recurrence risk of a recessive disorder occurring in the family. On a broader level, information on disease frequency can be used to plan diagnostic and therapeutic facilities and/or to organise targeted screening for specific populations. However, the necessary information on disease frequency is often lacking, while frequencies tend to vary widely across populations (Reich and Lander, 2001).

Recent papers have presented a method to estimate the total pathogenic allele frequency for an autosomal recessive disorder in a certain population by using information on the proportions of homozygous or compound heterozygous mutations among the offspring of consanguineous couples (Ten Kate et al., 2010; Gialluisi et al., 2012). As the proportions of homozygous and compound heterozygous cases depend on the total pathogenic allele frequency (q), the number and relative frequencies of pathogenic mutations, and the inbreeding coefficient (F) of the parents, q can be derived from knowledge of the other, here mentioned, parameters. The method has been applied by Gialluisi et al. for phenylketonuria, familial Mediterranean fever, and Wilson disease in various populations (2012; 2013). This approach, however, was shown to result in a biased estimation of q caused by the use of the average of the inbreeding coefficient in the total sample (Chapter 6). An alternative method based on the maximum likelihood method, which uses the same mutational data, avoids this bias and is presented in this chapter. Here, this estimation method is applied on data of patients with familial Mediterranean fever (FMF) from Tunisia and from Morocco. FMF is an autosomal recessive disorder caused by mutations in the *MEFV* gene, leading to short episodes of inflammation of serosal tissues, such as the peritoneum, synovia and pleura. Recurrent episodes with fever and abdominal pain are the most frequent clinical symptoms (Ben-Chetrit and Levy, 1998). In many patients that meet the clinical diagnosis of FMF, molecular testing fails to show two mutations. Although this can partly be explained by undiscovered recessive *MEFV* mutations, in addition, since extensive sequencing of the *MEFV* gene often does not identify two pathological mutations, it is suggested that etiological mechanisms other than bi-allelic mutations in the *MEFV* gene may play a role (Shohat and Halpern, 2011).

In this chapter we focus on autosomal recessive *MEFV*-FMF and aim to estimate the frequency of bi-allelic *MEFV* mutations causing FMF, by using the mutation data of 293 unrelated Tunisian and 199 unrelated Moroccan patients with clinical FMF. Since in many of these patients only one or no mutations were found, as is often the case in

clinically suspected FMF, we explore to what extent possible undiscovered recessive *MEFV* alleles may influence the estimate.

METHODS

Patients

For the Tunisian cohort, the estimates are based on the mutational data of 293 unrelated patients from all parts of Tunisia who underwent genetic testing for FMF between 2005 and 2010 in the molecular laboratory of the department of Hereditary and Congenital diseases at Charles Nicole Hospital in Tunis. The patients were referred by their physicians to the department based on definite or probable FMF according to Tel Hashomer clinical criteria (Livneh et al., 1997). Molecular genetic testing consisted of sequencing exon 10, and in case this failed to show two mutations, subsequently, it was followed by sequencing exon 2 of the *MEFV* gene. If only one mutation was identified, direct sequencing of the remaining eight exons was performed (Table 1a). Data on 131 of these patients were previously published (Chaabouni et al., 2007). An inbreeding coefficient was calculated for every patient, based on information from the patient or patient's parents (Table 1b).

Table 1a. Identified mutations in a cohort of Tunisian FMF patients

Homo-/Heterozygosity	Genotype	Number
Homozygous (total = 67)	c.2080A>G (p.(M694V))/c.2080A>G (p.(M694V))	22
	c.2040G>C/A (p.(M680I))/c.2040G>C/A (p.(M680I))	20
	c.2082G>A (p.(M694I))/c.2082G>A (p.(M694I))	19
	c.442G>C (p.(E148Q))/c.442G>C (p.(E148Q))	6
Compound heterozygous (total = 25)	c.2080A>G (p.(M694V))/c.2082G>A (p.(M694I))	5
	c.442G>C (p.(E148Q))/c.2040G>C/A (p.(M680I))	3
	c.2080A>G (p.(M694V))/c.2084A>G (p.(K695R))	2
	c.442G>C (p.(E148Q))/c.2082G>A (p.(M694I))	2
	c.2082G>A (p.(M694I))/c.2177T>C (p.(V726A))	1
	c.2080A>G (p.(M694V))/c.2177T>C (p.(V726A))	1
	c.2082G>A (p.(M694I))/c.2230G>T (p.(A744S))	1
	c.2080A>G (p.(M694V))/c.2076_2078del (p.(I692del))	1
	c.2076_2078del (p.(I692del))/c.2177T>C (p.(V726A))	1
	c.442G>C (p.(E148Q))/c.2230G>T (p.(A744S))	1
	c.442G>C (p.(E148Q))/c.2076_2078del (p.(I692del))	1
	c.442G>C (p.(E148Q))/c.2080A>G (p.(M694V))	1
	c.442G>C (p.(E148Q))/c.2177T>C (p.(V726A))	1

Homo-/Heterozygosity	Genotype	Number
	c.442G>C (p.(E148Q))/c.329T>C (p.(L110P))	1
	c.442G>C (p.(E148Q))/c.605G>A (p.(R202Q))	1
	c.2040G>C/A (p.(M680I))/c.2080A>G (p.(M694V))	1
	c.2040G>C/A (p.(M680I))/c.2177T>C (p.(V726A))	1
One mutation (total = 34)	c.442G>C (p.(E148Q))/N	13
	c.2080A>G (p.(M694V))/N	4
	c.2082G>A (p.(M694I))/N	4
	c.2040G>C/A (p.(M680I))/N	3
	c.2230G>T (p.(A744S))/N	3
	c.2177T>C (p.(V726A))/N	3
	c.2038A>C (p.(M680L))/N	1
	c.2084A>G (p.(K695R))/N	1
	c.2160C>G (p.(I720M))/N	1
	c.586G>T (p.(G196W))/N	1
No identified mutations		167
Total		293

N = not-identified

Table 1b. Tunisian cases by inbreeding coefficient

F	Number of cases
0.1250	1
0.0625	69
0.0313	17
0.0156	9
0.0078	12
0.0039	2
0.0000	183
Total	293

The Moroccan cohort consisted of 199 unrelated patients from all parts of Morocco who were referred to the department of Medical Genetics in Rabat in the period 2000-2010 and for whom molecular diagnostics was performed at the sole genetic laboratory that performs genetic testing for FMF in Morocco. All patients were referred by their physicians because of symptomatic criteria of FMF, ranging from probable to definite FMF, according to the Tel Hashomer criteria. Exons 2 and 10 of the *MEFV* gene were amplified by polymerase chain reaction and screened for mutations through direct sequencing. Data on 120 of these Moroccan patients were previously published (Belmahi et al., 2012) (Table 2a). Inbreeding coefficients were calculated based on pedigree data as reported by the patient or patient's parents (Table 2b).

It can be assumed that among the patients in our sample with only one or no identified mutations, cases are present in which the mutation in the *MEFV* gene was missed but is nevertheless present, especially in the case of the Moroccan patients in our sample in whom only exons 2 and 10 were sequenced. In Morocco, in 75% of the patients only one or no alleles were identified. In Tunisia this percentage is 69%.

Table 2a. Identified mutations in a cohort of Moroccan FMF patients

Homo-/Heterozygosity	Genotype	Number
Homozygous (total = 32)	c.2080A>G (p.(M694V))/c.2080A>G (p.(M694V))	19
	c.2082G>A (p.(M694I))/c.2082G>A (p.(M694I))	10
	c.2038A>C (p.(M680L))/c.2038A>C (p.(M680L))	2
	c.2076_2078del (p.(I692del))/c.2076_2078del (p.(I692del))	1
Compound heterozygous (total = 17)	c.2080A>G (p.(M694V))/c.2082G>A (p.(M694I))	9
	c.2080A>G (p.(M694V))/c.2230G>T (p.(A744S))	1
	c.2082G>A (p.(M694I))/c.442G>C (p.(E148Q))	1
	c.2080A>G (p.(M694V))/c.2081_2083del (p.(M694del))	1
	c.2082G>A (p.(M694I))/c.2081_2083del (p.(M694del))	2
	c.2080A>G (p.(M694V))/c.2282G>A (p.(R761H))	1
	c.2080A>G (p.(M694V))/c.2076_2078del (p.(I692del))	1
	c.2230G>T (p.(A744S))/c.2177T>C (p.(V726A))	1
One identified mutation (total = 32)	c.2080A>G (p.(M694V))/N	11
	c.2082G>A (p.(M694I))/N	7
	c.2076_2078del (p.(I692del))/N	2
	c.442G>C (p.(E148Q))/N	3
	c.2230G>T (p.(A744S))/N	7
	c.2282G>A (p.(R761H))/N	1
	c.2084A>G (p.(K695R))/N	1
No identified mutations		118
Total		199

N = not-identified

Table 2b. Moroccan cases by inbreeding coefficient

F	Number of cases
0.0625	30
0.0313	4
0.0156	7
0.0078	1
0.0039	4
0.0000	129
Total	175

Maximum Likelihood Estimator

As described in Chapter 6 the maximum likelihood estimator (MLE) for estimating the total pathogenic allele frequency assumes that, for every individual in the data set, two alleles have been observed. Since this is not the case in the Tunisian and Moroccan data, we propose applying two methods to account for this, leading to two estimates for each country:

Method 1. The allele frequency is estimated based on the data of all affected individuals with two observed pathogenic alleles. This yields an estimate that tends to be, in general, too low, since unknown, pathogenic alleles are ignored as disease alleles.

Method 2. One fictive allele is modelled to account for all missing alleles in FMF patients; all unobserved alleles are assigned to this fictive allele. By this assumption, for all affected individuals two alleles are now "observed" and the maximum likelihood estimator can be applied. The estimate that is found for q tends to be, in general, an overestimate, since not all unknown alleles in FMF patients have to be pathogenic.

Confidence intervals for q were constructed as was described in Chapter 6.

RESULTS

We computed the two estimates based on the Tunisian data and the Moroccan data.

FMF Tunisia. By applying method 1, the allele frequency q for autosomal recessive *MEFV*-FMF in Tunisia was estimated as 0.01040 (95% confidence interval (CI):0.00381-0.02540). When all missing alleles were assigned to the fictive allele (method 2) the estimate for q did not change substantially: 0.00949 (95% CI: 0.00460-0.01831). In this case, the second method did not result in a higher estimation, being the opposite of what was *a priori* expected.

FMF Morocco. When leaving out the individuals with one or two missing alleles (method 1), q was estimated as 0.00000 (95% CI: 0.00000-0.01326). The estimate equalled 0.0000 because the bi-allelic patients of consanguineous parents were all homozygous, and a positive proportion of compound heterozygous patients is needed to find a positive estimate of q. When the patients with the missing alleles were included in the data set by including one fictive allele in the model that accounts for all missing alleles (method 2), q was estimated as 0.002866 (95% CI: 0.00042-0.0118).

DISCUSSION

In this study we found that, by applying two methods – a method based on only the observed bi-allelic mutations as well as a method that accounts also for non-identified

alleles – the maximum likelihood method allows us to arrive at meaningful estimates and confidence intervals for *q*.

In many high-risk populations, the allele frequencies are estimated to be up to 1 in 5 and in some even up to 1 in 3 (Touitou, 2001; Stoffman et al., 2000; Yilmaz et al., 2001). Although one Algerian study from 2011, which tested 230 healthy controls from blood transfusion centres in the capital city, estimated a pathogenic *MEFV* allele frequency of approximately 0.1 (Ait-Idir et al., 2011), the carrier frequency in the Maghreb population (i.e. Morocco, Tunisia, Algeria) was estimated earlier, from a study that tested 113 healthy unrelated individuals, to be less than 1% (Belmahi et al., 2006). Our current results are consistent with this latter estimate. Among the Moroccan patients, only a few patients with bi-allelic homozygous mutations in combination with reported consanguineous parents were present that could contribute to the estimation. The estimate for Tunisia should therefore be considered more robust. The fact that the first method did not result in a lower estimate than the second method for the Tunisian data can be explained by stochastic variation in the estimate.

Many factors can influence the estimation of the pathogenic allele frequency when performed this way. The mutation analyses performed in the two different sets of patients were different from each other. While for the Tunisian patients who were found to be heterozygous, all other exons were sequenced, for the Moroccan patients the analysis was restricted to a targeted mutation analysis and sequence analysis of exons 2 and 10. The number of additional discovered alleles in the Tunisian sample is not substantial, however. It can be expected that in both data sets there were some undiscovered recessive *MEFV* alleles, especially those outside exons 2 and 10 and/or intronic mutations. However, as is described for other patient cohorts, given the large proportions of cases with clinical FMF, but no bi-allelic mutations, it can be concluded that in a large proportion of cases another etiological mechanism must be underlying the FMF phenotype (Shohat and Halpern, 2011).

For all calculations, the inbreeding coefficient is one of the variables. We used what was known about the value for *F* without making any further effort to obtain more precise estimates by interrogation or DNA analysis. Among the Tunisian individuals with *F*=0, who were only used for estimating the relative allele frequencies, 60% had homozygous mutations, whereas based on the estimated allele frequencies from people with a known *F,* this percentage was expected to be approximately 23%. A possible explanation is that the pedigree inbreeding coefficient that was calculated based on the reported ancestry is in fact an underestimation of the real genomic inbreeding coefficient. In communities with a tradition of consanguineous marriage, the estimation of *F* based on the reported relationship is usually an underestimation due to hidden consanguineous loops (Liu et al., 2006). As a result, the cumulative coefficient of inbreeding will be higher than the value calculated for one, two or three generations. This phenomenon can result in an underestimation of *q,* because more homozygotes would be encountered at the expense of the proportion of compound heterozygotes.

Meanwhile, the value for the inbreeding coefficient can also be overestimated. For instance, in the paper of Gialluisi et al. the calculation of *F* was approached by attributing a general *F* to the seemingly unrelated couples, which was based on the percentages of first- and second cousin marriages in the area of origin (2013). In this scenario, the couples who report their consanguinity or whose consanguinity can be traced, are included as such in the calculation, while couples who claim not to be related are still attributed a value for *F* that may be discordant with the actual value. This may result in an overestimation of the inbreeding coefficient. In short, an exact inbreeding coefficient is difficult to estimate based on genealogical data, and might be best approached by observations based on genomic homozygosity data.

Another factor that might influence the estimation of the actual allele frequency is the fact that this – indirect – method depends on symptomatic patients. Homozygous FMF mutation carriers can remain asymptomatic, as was found for example in a Turkish study where 2 out of 49 healthy control individuals had bi-allelic mutations in the *MEFV* gene (Tunca et al., 2002). Incomplete penetrance of mutations in the *MEFV* gene might therefore have resulted in a less accurate estimation.

Nevertheless, by using the MLE method in practice, the total allele frequency of bi-allelic *MEFV* FMF can be estimated – provided that sufficient information on both molecular and genealogical data are available – even where it can be assumed that not all mutations were identified.