



White Paper 23-11

Estimating Carrier Frequency, Carrier Detection Rate, and Post-Test Carrier Risk for Recessive Disorders

Authors:

Shirley Wu

Created: September 28, 2015

Version: 1

Summary:

23andMe Carrier Status reports provide estimates of the chances of still being a carrier for people who do not have the variant(s) tested for a given recessive condition. This is known as the residual or post-test carrier risk. These estimates are population-specific and are based on the carrier frequency and carrier detection rate of the test in that population. Our reports provide post-test carrier risk estimates only for populations that have both types of data available.

This white paper describes our methodology for calculating post-test carrier risks, and for determining carrier frequency and carrier detection rates for the purposes of that calculation.

Definitions

Carrier frequency (pre-test carrier risk)

The carrier frequency is the expected probability of having a single disease-causing variant for an autosomal recessive condition in a given population.

Carrier detection rate (coverage)

Carrier detection rate is the percentage of carriers that would be detected by the test. In the context of carrier testing, this is equivalent to the proportion of disease-causing variants that would be detected by the test.

Post-test carrier risk (residual risk)

The post-test carrier risk is the probability of being a carrier given a negative result for the variants tested.

Determining Carrier Frequency

When available, population-specific carrier frequencies are taken directly from the scientific literature, with preference given to those published as part of professional guidelines such as those from the American College of Medical Genetics (ACMG). In cases where a published estimate is not available, we may derive carrier frequency (x) using one of the following methods:

1. From literature-derived population-specific disease incidence data using the Hardy-Weinberg equation, assuming Hardy-Weinberg equilibrium to be true, $x = 2\sqrt{J}(1 - \sqrt{J})$, where J is the reported incidence of the condition. Prevalence may not be substituted for incidence. Different methods should be considered if incidence is not believed to reflect the true carrier frequency (for example, in populations where carrier screening is routine and has reduced the disease incidence, or if expected variant combinations are not observed, or if the disease is believed to be significantly under-diagnosed due to low penetrance or lack of diagnostic testing).
2. By dividing the frequency of carriers for one or more variants by the estimated carrier detection rate represented by those variants in a specific population. This method may be used when incidence may not accurately reflect the true carrier frequency. This method should only be used when reliable estimates exist for both allele carrier frequencies and carrier detection rate. In cases where sources report the frequency of the variant(s)

(*e.g.* 1 in 200 chromosomes), the carrier frequency of those variants is twice that value (*e.g.* 1 in 100 people).

These methods represent possible approaches; however, the use of any of these methods is subject to the validity of the underlying source data and scientific discretion. Additional methods may also be appropriate; these would be assessed on a case-by-case basis.

When multiple approaches are possible, estimates should be compared and evaluated for consistency. Estimates that differ by more than 20% are considered unreliable and in these situations post-test carrier risk should not be provided. If the estimates differ by less than 20%, we use the estimate for the carrier frequency that would lead to a higher reported risk.

Determining Carrier Detection Rate

As with carrier frequency, carrier detection rate is population-specific. For instance, if the test consists of one variant that accounts for 99% of the disease-causing variants in people of European descent, the carrier detection rate of the test would be 99% in people of European descent.

When available, population-specific carrier detection rates are taken directly from the scientific literature, with preference given to those published as part of professional guidelines such as those from ACMG.

If the set of variants described by a guideline differs from our test, we reduce the carrier detection rate from the guideline by the amount represented by the variant(s) in question. The reduction is based on individual variant frequency estimates in affected patients provided by that guideline when available, and on published literature estimates otherwise.

In cases where no professional guidelines exist, we may derive carrier detection rate using one of the following methods:

1. By combining individual variant-specific carrier detection rates from the literature. The proportion of disease-causing variants attributable to individual variants reported in the literature may also be combined to obtain this value.
2. By counting the proportion of disease-causing variants in carriers or patients of the relevant ethnicity represented by the variants included in our test. If

multiple studies with individuals of relevant ethnicities exist, we may use a single large study, or we may pool individuals from multiple smaller studies together.

- a. Care should be taken to ensure that individuals counted across studies are unique and unrelated.
- b. Patient studies should use a method of identifying patients that is reasonably unbiased with regards to the variants they may have. For example, studies with patients diagnosed on the basis of symptoms or biochemistry and then sequenced or genotyped are appropriate for contributing to a carrier detection rate estimate, but studies that selected patients based on genotype may not be appropriate as they represent a biased sampling of the patient population.

These methods represent possible approaches; however, the use of any of these methods is subject to the validity of the underlying source data and scientific discretion. Additional methods may also be appropriate; these would be assessed on a case-by-case basis.

Estimating Post-Test Carrier Risk

If a reliable published estimate for the post-test carrier risk is available for the exact set of variants used in the 23andMe test, we use the published estimate. Otherwise, for populations for which we can estimate carrier frequency and carrier detection rate as described above, we use the Bayesian method published by Ogino et al. (2004) to calculate the post-test risk of being a carrier:

$$p(\text{carrier} \mid \text{negative test}) = \frac{(1 - c)x}{1 - xc}$$

where x is the pre-test probability of being a carrier (equivalent to the carrier frequency), and c is the carrier detection rate of the test. These calculations make the assumption that the false positive rate of the test is $\ll cx$ and the false negative rate of the test is $\ll (1 - c)$.

Post-test carrier risks provided in our reports may be rounded for simplicity.

Acknowledgments

Erynn S. Gordon, Sheel Dandekar, Joyce Tung, Bethann Hromatka, and other members of the Product Science, Research, and Medical Affairs teams provided helpful feedback on this article.

References

Ogino S, Wilson RB. (2004) Bayesian analysis and risk assessment in genetic counseling and testing. *J Mol. Diagn.* 6:1-9.