



## Hardy Weinberg Equilibrium Data

Created: July 7, 2005; Updated: February 18, 2014.

**I am having trouble understanding how the Hardy-Weinberg Probability (HWP) value is calculated for SNP records. I'm no statistician, so it would be helpful if you could include an example.**

A good explanation of HWP and how it is calculated can be found in an online [Wikipedia article](#).

Currently, we use the chi-square and p-value in our HWP calculations. Here is an example of how we calculate the HWP for *ss221* in a population (population id as 506) following the HWP equation.

GtyFreq	subsnp_id	pop_id	gty_str	cnt
0.411	221	506	A/G	37.000000
0.144	221	506	A/A	13.000000
0.444	221	506	G/G	40.000000

cnt	AlleleFreq	subsnp_id	pop_id	allele
117.000000	0.650	221	506	G
63.000000	0.350	221	506	A

subsnp_id	pop_id	ind_cnt
221	506	90.000000

subsnp_id	pop_id	chr_cnt
221	506	180.000000

subsnp_id	pop_id	Degree of Freedom
221	506	1

exp_gtyFreq	subsnp_id	pop_id	exp_gtyCnt	allele_1	allele_2	gty_str
0.422	221	506	38.025	G	G	G/G
0.455	221	506	40.950	G	A	A/G
0.122	221	506	11.025	A	A	A/A

subsnp_id	pop_id	gty_str	observed_gtyCnt
221	506	G/G	40.00
221	506	A/G	37.00
221	506	A/A	13.00

subsnp_id	pop_id	Chi-square
221	506	0.837

After obtaining the chi-square value, calculate the p-value using the degrees of freedom via the gamma function, or many websites offer [p-value calculators](#). (9/11/07)

**How is HWP (Hardy Weinberg Probability) found in the Population Diversity section of the refSNP report determined, and how does it relate to genotype data?**

A good explanation of HWP and how it is calculated can be found [online](#). The next question in this section also contains information you should find useful. (6/4/07)

**Your Hardy Weinberg probabilities are calculated using 2 df—an incorrect calculation. Because only one independent parameter is being used to estimate ( $p=1-q$ ), = there should only be 1 df.**

Thanks for pointing out the inclusion of failed assays in the Hardy Weinberg estimates for refSNP (rs) clusters appearing on rs cluster reports as well as in the genotype and allele frequency reports. We will exclude such genotypes in the Hardy Weinberg equilibrium calculations for the release of dbSNP build 120.

Many of the assumptions of the Hardy Weinberg equilibrium are not necessarily met in the submissions to the dbSNP, but we do believe that there is utility (although limited) in calculation of Hardy Weinberg equilibrium over rs clusters. Please see the Hardy Weinberg equilibrium performed over submitted SNPs (ss) by population. These are listed in the submitted SNP details page as well as in the genotype and allele frequency report. Please be aware, however, that the population in the context of dbSNP may (or may not) differ from a population as defined in a population genetics context.

**The displayed value for HWP on your web page is different than the FAQ which calculates it. Am I using the correct calculation method?**

To speed up database update, we have used a lookup table which is binned. This is a contributory factor as to why the P value is not exact but should be close. To calculate the P value (chi-square distribution), we used the Gamma Function. I think we got the C algorithm from Numeric Recipe book.

If you are interested in HWE, have you looked at Fisher Exact test? From what biologists have told us, the Fisher Exact test is a better way to estimate HWP especially when sample size is small. When time allows, we hope to switch our HWE calculations from the Chi-square test to the Fisher Exact test.(08/01/08)