



Johns Hopkins University, Dept. of Biostatistics Working Papers

9-13-2004

Effect of Misreported Family History on Mendelian Mutation Prediction Models

Hormuzd A. Katki

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics & Div. of Cancer Epidemiology & Genetics, National Cancer Institute, katkih@mail.nih.gov

Suggested Citation

Katki, Hormuzd A., "Effect of Misreported Family History on Mendelian Mutation Prediction Models" (September 2004). *Johns Hopkins University, Dept. of Biostatistics Working Papers*. Working Paper 58.
<http://biostats.bepress.com/jhubiostat/paper58>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

Effect of Misreported Family History on Mendelian Mutation Prediction Models

Hormuzd A. Katki

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD 21205, U.S.A.

and

Division of Cancer Epidemiology and Genetics, NCI, NIH, DHHS
6120 Executive Blvd. Room 8044 Rockville, MD 20852, U.S.A.

September 13, 2004

SUMMARY. People with familial history of disease often consult with genetic counselors about their chance of carrying mutations that increase disease risk. To aid them, genetic counselors use Mendelian models that predict whether the person carries deleterious mutations based on their reported family history. Such models rely on accurate reporting of each member's diagnosis and age of diagnosis, but this information may be inaccurate. Commonly encountered errors in family history can significantly distort predictions, and thus can alter the clinical management of people undergoing counselling, screening, or genetic testing. We derive general results about the distortion in the carrier probability estimate caused by misreported diagnoses in relatives. We show that the Bayes Factor that channels all family history information has

email: katkih@mail.nih.gov

a convenient and intuitive interpretation. We focus on the ratio of the carrier odds given correct diagnosis vs. given misreported diagnosis to measure the impact of errors. We derive the general form of this ratio and approximate it in realistic cases. Misreported age of diagnosis usually causes less distortion than misreported diagnosis. This is the first systematic quantitative assessment of the effect of misreported family history on mutation prediction. We apply the results to the BRCAPRO model, which predicts the risk of carrying a mutation in the breast and ovarian cancer genes BRCA1 and BRCA2.

KEY WORDS: Family History Errors; Genetic Counselling; Bayes Factor; Measurement Error; BRCA1; BRCA2; BRCAPRO; CRCAPRO

1. Introduction

People who are concerned that their family has a high prevalence of disease may seek genetic counselling to assess their risk of carrying inherited genetic mutations that cause the disease (Croyle and Lerman, 1999). To aid such people (referred to as consultands), genetic counselors employ statistical models that predict whether the consultand carries deleterious mutations by using the consultand's reported family history of disease. For syndromes whose onset occurs over a lifetime (as is common in cancer), family history is the age at which each family member developed disease, or that member's current age or age at death. Models that combine family history information with knowledge of each gene's mode of Mendelian transmission, penetrance (probability of disease given genotype), and prevalence are called Mendelian models. These models determine the probability that the consultand is a mutation carrier via Bayes's rule and Mendel's laws (Murphy and Mutalik, 1969). The consultand's carrier probability is a crucial component in the

consultand's decision to take a genetic test (if a test exists), to undergo frequent disease screening, or to consider prophylactic options. For the Breast-Ovarian Cancer Syndrome (Claus et al., 1996), the popular Mendelian model BRCAPRO estimates the probability that a consultand carries a deleterious mutation in the BRCA1 and BRCA2 genes, based on family history of breast and ovarian cancer (see Berry et al., 1997 and Parmigiani et al., 1998). Another example is CRCAPRO, which computes the probability of carrying a mutation in the genes MLH1 and MSH2 given family history of colorectal and endometrial cancer (Chen et al., 2004).

However, Mendelian models rely on accurate knowledge of family history. Consultands cannot always provide accurate information; sometimes they cannot provide the required information, or mistakenly provide inaccurate information. Error rates depend on many variables, including disease type, age-at-diagnosis, and degree of relationship to the consultand; for example, consultands report only 44% of the ovarian cancers in their second-degree (2°) relatives (Ziogas and Anton-Culver, 2003).

Such errors can seriously distort the carrier probability estimate. For example, consider an Ashkenazi-Jewish consultand reporting herself as cancer-free at age 50, reporting that her mother got breast cancer at age 70, but unaware that her grandmother had ovarian cancer at age 50. Then BRCAPRO estimates her probability of carrying a BRCA mutation as 3.5%, but her BRCAPRO probability based on correct information is 19%. This is especially important because genetic counselors often offer genetic testing to the consultand once the probability breeches 10% (Domchek et al., 2003).

Genetic counselors working in academia or controlled studies try to con-

tact other relatives to verify the given information, but relatives may be deceased or otherwise difficult to contact. Furthermore, genetic counselors not working in those settings may have limited options for verifying the given information. Thus inaccurate family history is a reality of genetic counseling, but we are unaware of any methodological contribution addressing this complex problem. This paper examines Mendelian models to determine the distortion of the carrier probability by three types of misreported family history:

1. Diagnosis incorrect, Age-at-diagnosis correct: This error results from the consultand knowing that something happened to their relative at that age, but not knowing what. In the example above, ovarian cancer diagnoses are often concealed from the family. Since ovarian cancer is often rapidly fatal, the age-at-death approximates the true age-at-diagnosis of ovarian cancer. This paper primarily concerns this error.
2. Diagnosis correct, Age-at-diagnosis incorrect: Here the consultand remembers a relative's disease diagnosis but not their age at diagnosis. This error also happens if the model requires age-of-onset rather than age-of-diagnosis. Often, all that is known is an age range within which the relative developed disease.
3. Diagnosis incorrect, Age-at-diagnosis incorrect.

This paper shows a convenient interpretation of the Bayes Factor (BF) from Mendelian models. Defining the distortion of the estimated carrier odds caused by errors as the ratio of the BF given correct information to the BF given misreported information, we derive the distortion caused by the first

type of error in each family member, both exactly and for a rare autosomal-dominant mutation, for a general Mendelian model and for BRCAPRO. This distortion takes the form of a hazard ratio. Misreported diagnoses usually cause more distortion than misreported ages, so this paper focuses on this aspect. The last two types of error are considered within BRCAPRO.

2. Computing the Carrier Probability

Mendelian models require knowledge of which disease each relative developed and the age when it was diagnosed, or if no disease developed, the current age or age of death. For simplicity, assume that only one disease is involved in the syndrome, so $c_i = 0, 1$ indicates disease in relative i and y_i is the age of diagnosis (or current age or age at death). Let $H_i = (y_i, c_i)$ and $H = (H_0, H_1, \dots)$. Assume that individuals independently inherit one allele from each parent at each autosomal locus and that the alleles are either normal or mutated. Let $\gamma_i = 0, 1$ indicate carrying the genotype(s) that confer(s) disease risk; for example, $\gamma_i = 1$ for a dominant trait when the member carries at least 1 mutant allele, but for a recessive trait $\gamma_i = 1$ implies the relative carries two mutant alleles. The absolutely continuous penetrance survivals S_0, S_1 are the probability of surviving the disease up to some age, given $\gamma_i = 0, 1$ respectively; similarly for the penetrance densities f_0, f_1 and penetrance hazards h_0, h_1 . The population prevalence of $\gamma_i = 1$ is π . The aim is the consultand's carrier probability $P(\gamma_0 = 1|H)$.

By Bayes rule, the odds of the consultand being a carrier is a product of

the prior odds in the population and the Bayes Factor:

$$\frac{P(\gamma_0 = 1|H)}{P(\gamma_0 = 0|H)} = \frac{\pi}{1 - \pi} BF(H),$$

$$BF(H) = \frac{P(H|\gamma_0 = 1)}{P(H|\gamma_0 = 0)}.$$

Since family history only affects the carrier probability through the Bayes Factor (BF), it suffices to consider the effect of errors in family history on the BF. Section 2.1 shows and interprets the BF, and later sections examine the effect of errors on the BF.

2.1 Bayes Factor for a General Family

If all relationships between the ancestors and the descendants are via the consultand, then conditioning on the consultand makes the ancestors independent of the descendants. The BF breaks into contributions from the (c)onsultand, the (a)ncestors, and the (d)escendants. Thus $BF(H) = BF_c \times BF_a \times BF_d$.

For the consultand, the likelihood is

$$P(H|\gamma_0) = P(Y_0 = y_0, C_0 = c_0|\gamma_0) = P(H_0|\gamma_0) \propto f_{\gamma_0}(y_0)^{c_0} S_{\gamma_0}(y_0)^{1-c_0},$$

assuming independent non-informative censoring. The censoring mechanism always drops out of the likelihood and will never be referred to (Katki et al., 2004). The BF is

$$BF_c = \frac{P(H_0|\gamma_0 = 1)}{P(H_0|\gamma_0 = 0)} = \frac{f_1(y_0)^{c_0} S_1(y_0)^{1-c_0}}{f_0(y_0)^{c_0} S_0(y_0)^{1-c_0}}. \quad (1)$$

BF_c is a ratio of penetrance densities if the consultand has the disease, otherwise, it is the ratio of penetrance survivals.

Assuming that H_i is independent of all other information, given that member's γ_i , then contributions to the BF from family members are channeled through the first-degree (1°) relatives, i.e. parents and children. Thus BF_a is funneled through the parents, so conditioning on their genotypes breaks the ancestors into independent contributions from mother's side ancestors, father's side ancestors and siblings' families (who contribute only through their connection to the parents). The likelihood is

$$\sum_{i,j=0}^1 \left(\prod_{s=1}^{\#sibs} P(H_s | \gamma_f = i, \gamma_m = j) \right) P(H_f | \gamma_f = i) P(H_m | \gamma_m = j) P(\gamma_f = i, \gamma_m = j | \gamma_0), \quad (2)$$

where H_f is the vector of all the father's side phenotypes, H_m is the mother's side phenotypes and H_s are each sibling's side phenotypes. Let A_{ij} be each term in (2) when $\gamma_0 = 0$ and $w_{ij} = A_{00}A_{01}A_{10}A_{11}/A_{ij}$. Then

$$BF_a = \frac{w_{00}^{-1}B_{00} + w_{01}^{-1}B_{01} + w_{10}^{-1}B_{10} + w_{11}^{-1}B_{11}}{w_{00}^{-1} + w_{01}^{-1} + w_{10}^{-1} + w_{11}^{-1}},$$

where B_{ij} is the BF contributed by the parents if their genotypes were known:

$$B_{ij} = \frac{P(\gamma_f = i, \gamma_m = j | \gamma_0 = 1)}{P(\gamma_f = i, \gamma_m = j | \gamma_0 = 0)}.$$

Thus BF_a is a weighted average of BFs that the parents would contribute if their genotypes were known. Since their genotypes are unknown, the BF is a weighted average of the two, with weights w_{ij} reflecting the likelihood of the parents' possible genotype. However, if backwards transmission of $\gamma_0 = 0$ is impossible, then one of the $B_{ij} = 1/0 = \infty$ and the weighted-average interpretation breaks down.

The descendants are the kids of the consultand and their families. Their likelihood contribution is

$$\prod_{k=1}^{\#kids} \sum_{i=0}^1 P(H_k | \gamma_k = i) P(\gamma_k = i | \gamma_0)$$

where H_k are all the phenotypes within the family of kid k . Let $w_i(k) = P(H_k|\gamma_k = i)P(\gamma_k = i|\gamma_0 = 0)$. Then

$$BF_d = \prod_{k=1}^{\#kids} \frac{w_0^{-1}(k)B_0 + w_1^{-1}(k)B_1}{w_0^{-1}(k) + w_1^{-1}(k)},$$

where $B_i = p_{1i}/p_{0i} = P(\gamma_k = i|\gamma_0 = 1)/P(\gamma_k = i|\gamma_0 = 0)$ are the Bayes Factors favoring the consultand being a carrier, if the child's carrier status is known. Thus BF_d is a weighted averaged of the BFs that children would contribute if their genotypes are known, with weights reflecting the likelihood of each child's possible genotype.

Thus the BF breaks into pieces from the consultand and family. The family contributes a weighted average of the BFs from 1° relatives if their genotypes were known. For simple families, such as consultand/mother or consultand/mother/maternal grandmother, the weights are the likelihood of the relatives' data given the consultand's carrier status and the BFs that are averaged over are the likelihoods of the mother being a carrier given the consultand's carrier status – see the Appendix, which also provides approximate BFs for a rare autosomal-dominant mutation. The rest of the paper shows the distortion of BF by the three types of errors of section 1.

3. Effect of Misreported Diagnosis

Under the first type of error (which assumes known age of diagnosis), the effect of misreported diagnosis on the BF can be summarized by a ratio of BFs, the numerator being the BF for the family member correctly having disease and the denominator for the family member incorrectly not having disease. This is an underreporting error; for overreporting, the ratio is inverted.

If the consultand has underreported cancer, then the BF ratio is

$$BF \text{ Ratio} = \frac{f_1/f_0}{S_1/S_0} = \frac{f_1/S_1}{f_0/S_0} = \frac{h_1}{h_0}, \quad (3)$$

the hazard ratio of the mutation vs. wildtype (since age of diagnosis is known, it is dropped as an argument). Thus errors in diagnosis are irrelevant if the diagnoses imply equivalent hazards at the age of diagnosis. Furthermore, if the hazards are proportional, then misreporting causes the same distortion at any age. Since the BF factors into contributions from the consultand and from the family, the BF ratio for the misreporting in the family is a product of (3) with the BF ratio from the family derived below. It is most useful to consider the effect of misreporting diagnosis in a single 1° or 2° relative.

3.1 Misreporting a 1° relative's diagnosis

In a consultand/1°-relative family with underreported 1° relative's diagnosis, then the ratio of BFs is (see Appendix for derivation of the BF)

$$BF \text{ Ratio} = \frac{f_0 p_{10} + f_1 p_{11}}{f_0 p_{00} + f_1 p_{01}} \times \frac{S_0 p_{00} + S_1 p_{01}}{S_0 p_{10} + S_1 p_{11}} = \frac{h_1^{(1)}}{h_0^{(1)}}, \quad (4)$$

$$h_j^{(1)} = \frac{(S_1 p_{j1})^{-1} h_0 + (S_0 p_{j0})^{-1} h_1}{(S_1 p_{j1})^{-1} + (S_0 p_{j0})^{-1}}. \quad (5)$$

The $h_j^{(1)}$ are weighted averages of h_0, h_1 . Since a ratio of weighted averages of two quantities is less than the ratio of the two quantities, then misreporting a 1° relative distorts the BF less than the same misreport in the consultand. Furthermore, the weighting depends on the inverse of $S_j p_{ij} = P(Y_1 = y_1, C_1 = 0, \gamma_1 = j | \gamma_0 = i)$. This is the likelihood of the 1° relative surviving, given the consultand. If this probability for j is small, then the other hazard h_{1-j} gets more weight. If the hazards are proportional, there is still time-dependence in equation (4) since the weights depend on

time through the survivals.

For a rare autosomal-dominant mutation (see Appendix equation A.1), the BF ratio is approximately

$$\frac{h_1^{(1)}}{h_0^{(1)}} \approx \frac{1 + \frac{f_1}{f_0}}{1 + \frac{S_1}{S_0}} = \frac{1}{S_0 + S_1} \left(S_0 + S_1 \frac{h_1}{h_0} \right) = \frac{1}{1 + \frac{S_1}{S_0}} \left(1 + \frac{S_1}{S_0} \frac{h_1}{h_0} \right). \quad (6)$$

Equations (6) and (3) are the same except for the '1+' terms, and (6) is a linear function of the hazard ratio, which is the effect of an underreport for the consultand. The intercept and slope sum to one, and each are posterior probabilities of being a non-carrier (and carrier, respectively) given being unaffected, assuming even prior odds of being a carrier. Thus, when $h_1/h_0 = 1$, then $h_1^{(1)}/h_0^{(1)} = 1$; no-distortion in the consultand is also no-distortion for the 1° relative. The distortion caused by misreported diagnosis in 1° relatives is less than that caused by the same misreported diagnosis in the consultand (figure 1). At early ages of diagnosis, when $S_0 \approx S_1 \approx 1$, the distortion in the BF by misreported diagnoses in 1° relatives is half that of the same misreported diagnosis in the consultand. Furthermore, if the mutation is strongly deleterious and shifts the age of onset to earlier ages (like BRCA), then the survival ratio is small at older ages and the hazard ratio decreases at older ages, and the distortion attenuates. Thus misreported diagnoses in old enough 1° relatives cause little distortion. Intuitively, if $S_1(y) \approx 0$, then the relative has reached an age that carriers rarely reach, thus the relative is unlikely to be a carrier. It doesn't matter what happens at age y , just reaching that age suffices.

[Figure 1 about here.]

3.2 Misreporting a 2° Relative's Cancer

For the consultand/1°-relative/2°-relative family in the Appendix, the BF ratios, depending on the 1° relative's status c_1 , are $h_1^{(2)}/h_0^{(2)}$ where

$$h_i^{(2)} = \frac{(f_1^{c_1} S_1^{1-c_1} S_1 p_{i1})^{-1} h_0 + (f_0^{c_1} S_0^{1-c_1} S_0 p_{i0})^{-1} h_1}{(f_1^{c_1} S_1^{1-c_1} S_1 p_{i1})^{-1} + (f_0^{c_1} S_0^{1-c_1} S_0 p_{i0})^{-1}}.$$

Again, this is a ratio of weighted averages of the hazards. Under the rare autosomal-dominant mutation approximation of (A.2), the ratios are

$$BF_{ratio}^1 = \frac{BF(c_1 = 1, c_2 = 1)}{BF(c_1 = 1, c_2 = 0)} \approx \frac{1 + \frac{f_1}{f_0} \left(1 + \frac{f_1}{f_0}\right)}{1 + \frac{f_1}{f_0} \left(1 + \frac{S_1}{S_0}\right)} \quad (7)$$

$$BF_{ratio}^0 = \frac{BF(c_1 = 0, c_2 = 1)}{BF(c_1 = 0, c_2 = 0)} \approx \frac{1 + \frac{S_1}{S_0} \left(1 + \frac{f_1}{f_0}\right)}{1 + \frac{S_1}{S_0} \left(1 + \frac{S_1}{S_0}\right)}. \quad (8)$$

These ratios are of the form $g(\alpha_1) = (1 + \alpha_1(1 + A))/(1 + \alpha_1(1 + B))$ where α_1 is contributed by the 1° relative and $(1 + A)/(1 + B)$ is the ratio if the 1° relative were misreported (equation (6)). Since $1 \leq f(\alpha_1) \leq (1 + A)/(1 + B)$, the effect of misreported diagnosis in the 2° relative is attenuated from the effect that the same error in the 1° relative would have had. Also,

$$\frac{h_1^{(2)}}{h_0^{(2)}} \approx \frac{1 + \alpha_1 \left(1 + \frac{f_1}{f_0}\right)}{1 + \alpha_1 \left(1 + \frac{S_1}{S_0}\right)} = \frac{1}{1 + \alpha_1 \left(1 + \frac{S_1}{S_0}\right)} \times \alpha_1 \left(1 + \frac{S_1}{S_0}\right) \frac{h_1^{(1)}}{h_0^{(1)}},$$

a linear function of the BF ratio for the 1° relative. The intercept and slope are less than 1, so we again have attenuation. Plugging in equation (6),

$$\frac{h_1^{(2)}}{h_0^{(2)}} \approx \frac{1}{1 + \alpha_1 \left(1 + \frac{S_1}{S_0}\right)} \times \left(1 + \alpha_1 + \alpha_1 \frac{S_1}{S_0} \frac{h_1}{h_0}\right),$$

the BF ratio in the 2° relative is linear in the BF ratio for the consultand (Figure 1). Thus, since $g(\alpha_1)$ is monotone increasing, and if $f_1/f_0 > S_1/S_0$

(but small enough so the rare mutation approximation (A.2) still holds; reasonable for BRCA after age 35), then misreported diagnosis in the 2° relative causes greater distortion if the 1° relative is affected rather than unaffected. But the change in the distortion with age of the 2° relative strongly depends on the status and age of the 1° relative, so no simple rules can be stated.

3.3 Application to *BRCAPRO*

Figure 2 shows the BF ratio for underreported breast cancer diagnoses in family members in the simple families for the BRCA1 carrier probability in Ashkenazi Jews. The penetrances are from Chen, Iversen, et al., 2004. The consultand is 10 and 20 years younger than the 1° and 2° relative respectively. Each solid line is the exact BF ratio for the family and the dotted line is the rare mutation approximation. Underreporting the consultand leads to BF ratios of 20 at young ages and decreases to 7 at old ages (as noted in section 3, these BF ratios are the hazard ratios). Underreporting a single 1° relative leads to BF ratios that go from 7.5 to 2.5, and the lowest two lines show the effect of underreporting a 2° relative fixing the 1° relative. The BF ratios for each error are closely log-linear in age because the hazard ratios are closely log-linear in age. The approximations tend to slightly overestimate the effect of underreporting. The approximation tends to be rougher if the penetrance density ratio is high (see Appendix), as it is at younger ages. Also, the requirements of the approximation for 2° relatives are more stringent than for 1° relatives (see Appendix), so the quality of the approximation depends on the family. However, the approximation does correctly predict that BF ratio for 2° relatives with affected 1° relatives is higher than for unaffected 1° relatives. For a rare mutation like BRCA, an underreporting BF ratio of just

two will halve the carrier probability, so these errors can have big impact.

[Figure 2 about here.]

4. Effect of Errors in Age of Diagnosis

Errors in age of diagnosis can be summarized by ratios of BFs, the numerator with the BF given the true age and denominator the BF with the incorrectly reported age. Often a consultand cannot recall the exact age, but can specify a range within which the true age lies. This is like a round-off error. Although general results are difficult to obtain, it is easy to simulate such errors in a specific model, such as BRCA_{PRO}. For the simple families of the Appendix, the two left panels of figure 3 show the distortion in the BF caused by rounding the age of a single 1° relative to the nearest 30 (± 15 years) and 10 (± 5 years) and the two right panels of figure 3 show the effect of rounding the age of a single 2° relative to the nearest 30 and 10, assuming an affected 1° relative with known age of diagnosis at the true age of diagnosis of the 2° relative. Errors of ± 15 years cause a maximum distortion of 70% in the BF, and a maximum 20% distortion for ± 5 years. Similar computations show that the BF ratio for rounding age to the nearest 30 in unaffected relatives is no more than 1.15. Figure 4 shows the effect of simultaneous underreporting and rounding errors. Clearly, underreporting diagnosis is the more important error, but figure 3 shows that strong rounding error can cause meaningful distortion.

[Figure 3 about here.]

[Figure 4 about here.]

5. Discussion

The three types of misreported family history discussed in this paper can seriously distort the carrier probability estimate from Mendelian mutation prediction models. The distortion caused by misreported family history is channelled through the ratio by which it changes the true BF. The BF is the product of contributions from the consultand and the ancestors/descendants that are weighted averages of the BFs that parents and children would contribute if their genotypes were known, with weights depending on the likelihood of them carrying each genotype.

Under this special structure, the BF ratios for misdiagnosis in a relative (assuming known age of diagnosis) are ratios of weighted averages of the mutant and wildtype penetrance hazards. These ratios are generally complicated functions involving the entire family, but can be simplified in certain realistic cases. Assuming a rare autosomal-dominant mutation (like BRCA), the BF ratios in higher-degree relatives are linear in the BF ratio for the consultand and are attenuated towards one. In this case, the distortion in the BF by misreporting the diagnosis of 1° relatives at young ages is half that of underreporting the consultand, but at old enough ages, the distortion attenuates. If the penetrance density ratio is greater than the survival ratio (like BRCA), then the effect of this error in 2° relatives is stronger if the 1° relative is affected. If the penetrance hazards are proportional (approximately true for BRCA1 and ovarian cancer), then the distortion caused by misdiagnosis in the consultand is the same for any age of diagnosis, although the distortion caused by this error in relatives generally depends on age.

These results were applied to BRCAPRO. The amount of distortion de-

creases with age of relative, but also depends on the relationship to the consultand and outcomes in other relatives. Misreporting age of diagnosis in BRCAPRO for unaffected relatives has little effect. But the same error in affected relatives, while not as important as misreporting diagnosis, can have meaningful effects if the error is strong enough. For a rare mutation like BRCA, an underreporting BF ratio of just two will halve the carrier probability, and figures 2 and 4 show that the distortion is often much worse.

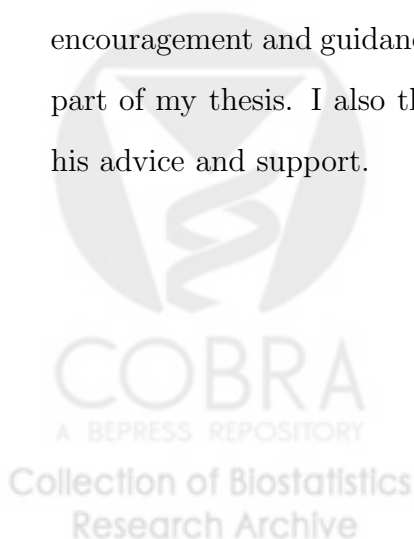
These results can alert genetic counselors to which types of reported family history require verification, for general Mendelian models and BRCAPRO in particular. Educating genetic counselors about these results is a major challenge. The intuition developed under the rare-mutation approximation could be useful for providing general guidelines. For a specific model, an easily interpretable version of the BF ratio graphs could be a handy reference for the counselor. The next step is proposing Mendelian models that handle misreported family history by accounting for population-based misreporting rates, such as those reported by Ziogas and Anton-Culver (2003).

ACKNOWLEDGEMENTS

I greatly thank my Ph.D thesis advisor, Giovanni Parmigiani, for his constant encouragement and guidance in this fruitful area of research, which comprises part of my thesis. I also thank my supervisor at NCI, Barry Graubard, for his advice and support.

RÉSUMÉ

REFERENCES



- Berry, D. A., Parmigiani, G., Sanchez, J., Schildkraut, J. and Winer, E. (1997). Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *Journal of the National Cancer Institute* **89**, 227–238.
- Chen, S., Iversen, E. S., Friebel, T., Finkelstein, D., Weber, B. L., Eisen, A., Peterson, L. E., Schildkraut, J. M., Isaacs, C., Peshkin, B. N., Corio, C., Leondaridis, L., Tomlinson, G., Amos, C. I., Strong, L. C., Berry, D. A., Dutson, D., Kerber, R., Euhus, D. M. and Parmigiani, G. (2004). Comprehensive evaluation of breast and ovarian cancer risks associated with BRCA1 and BRCA2 mutations. Submitted.
- Chen, S., Wang, W., Broman, K. W., Katki, H. A. and Parmigiani, G. (2004). BayesMendel: An R environment for Mendelian risk prediction. *Statistical Applications in Genetics and Molecular Biology* **3**. <http://www.bepress.com/sagmb/vol3/iss1/1>.
- Claus, E. B., Schildkraut, J. M., Thompson, W. D. and Risch, N. J. (1996). The genetic attributable risk of breast and ovarian cancer. *Cancer* **77**, 2318–2324.
- Croyle, R. T. and Lerman, C. (1999). Risk communication in genetic testing for cancer susceptibility. *Journal of the National Cancer Institute Monographs* **1999**, 59–66.
- Domchek, S. M., Eisen, A., Calzone, K., Stopfer, J., Blackwood, A. and Weber, B. L. (2003). Application of breast cancer risk prediction models in clinical practice. *Journal of Clinical Oncology* **21**, 593–601.
- Katki, H. A., Chen, S. and Parmigiani, G. (2004). Censoring and competing risks in Mendelian mutation prediction models. In Preparation.

- Murphy, E. A. and Mutalik, G. S. (1969). The application of Bayesian methods in genetic counselling. *Human Heredity* **19**, 126–151.
- Parmigiani, G., Berry, D. and Aguilar, O. (1998). Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *American Journal of Human Genetics* **62**, 145–158.
- Ziogas, A. and Anton-Culver, H. (2003). Validation of family history data in cancer family registries. *American Journal of Preventive Medicine* **24**, 190–198.

APPENDIX

Bayes Factors for Simple Families

Subscripts on H, γ denote degree of relationship to consultand. Consider a consultand-1° relative family. The likelihood is

$$P(H|\gamma_0) = P(H_0, H_1|\gamma_0, \gamma_1 = 0)P(\gamma_1 = 0|\gamma_0) + P(H_0, H_1|\gamma_0, \gamma_1 = 1)P(\gamma_1 = 1|\gamma_0).$$

Under the usual conditional independence assumptions,

$$P(H_0, H_1|\gamma_0, \gamma_1) = P(H_0|\gamma_0)P(H_1|\gamma_1) = f_{\gamma_0}(y_0)^{c_0} S_{\gamma_0}(y_0)^{1-c_0} f_{\gamma_1}(y_1)^{c_1} S_{\gamma_1}(y_1)^{1-c_1}.$$

The backwards transmission probabilities $p_{ij} = P(\gamma_1 = j|\gamma_0 = i)$ depend on the mode of Mendelian transmission. For example, if the mode is autosomal dominant, then $p_{01} = \pi/2$, $p_{11} = (1 + \pi)/2$. Putting these pieces together, the likelihood is

$$P(H|\gamma_0) = (f_{\gamma_0}(y_0)G(y_0))^{c_0}(S_{\gamma_0}(y_0)g(y_0))^{1-c_0} \times \\ ((f_0(y_1)G(y_1))^{c_1}(S_0(y_1)g(y_1))^{1-c_1}p_{\gamma_00} + (f_1(y_1)G(y_1))^{c_1}(S_1(y_1)g(y_1))^{1-c_1}p_{\gamma_01}),$$

and the BF is

$$BF(H) = BF_c \times \frac{M_{10} + M_{11}}{M_{00} + M_{01}} = BF_c \times \frac{\frac{p_{10}}{p_{00}} \frac{1}{M_{01}} + \frac{p_{11}}{p_{01}} \frac{1}{M_{00}}}{\frac{1}{M_{01}} + \frac{1}{M_{00}}}.$$

The BF breaks into a product of the BF from the consultand (BF_c) and a contribution from the 1° relative that is funneled through

$$M_{ij} = P(H_1, \gamma_1 = j | \gamma_0 = i) = P(H_1 | \gamma_1 = j) P(\gamma_1 = j | \gamma_0 = i).$$

Furthermore, the backwards transmission probability ratios $p_{1i}/p_{0i} = P(\gamma_1 = i | \gamma_0 = 1) / P(\gamma_1 = i | \gamma_0 = 0)$ are the Bayes Factors in favor of the consultand being a carrier, if the 1° relative's carrier status is known. Since it is unknown, the BF is a weighted average of the two, with weights M_{ij} reflecting how likely the 1° relative is a carrier. Assuming a rare autosomal-dominant mutation, the BF is approximately

$$BF \approx BF_c \times \frac{1}{2} \left(1 + \frac{f_1^{c_1} S_1^{1-c_1}}{f_0^{c_1} S_0^{1-c_1}} \right), \quad (\text{A.1})$$

assuming π is rare enough so that $\pi f_1 / f_0 \approx \pi S_1 / S_0 \approx 0$.

For a consultand/1° relative/2° relative family (where the 2° relative is directly related to the 1° relative, i.e. mother and maternal grandmother), using the usual conditional assumptions, the likelihood is

$$P(H | \gamma_0) = P(H_0 | \gamma_0) \times \sum_{\gamma_1, \gamma_2=0}^{\gamma_1, \gamma_2=1} P(H_1 | \gamma_1) P(H_2 | \gamma_2) P(\gamma_2 | \gamma_1) P(\gamma_1 | \gamma_0),$$

and the BF is

$$\begin{aligned} BF(H) &= \frac{P(H | \gamma_0 = 1)}{P(H | \gamma_0 = 0)} = BF_c \times \frac{M_{10} GM_0 + M_{11} GM_1}{M_{00} GM_0 + M_{01} GM_1} \\ &= BF_c \times \frac{\frac{p_{10}}{p_{00}} \frac{1}{M_{01} GM_1} + \frac{p_{11}}{p_{01}} \frac{1}{M_{00} GM_0}}{\frac{1}{M_{01} GM_1} + \frac{1}{M_{00} GM_0}}, \end{aligned}$$

where $GM_i = P(H_2|\gamma_1 = i)$. The family's contribution is a weighted average of the BFs if the 1° relative's carrier status is known. This is also the BF for a family with one 1° and one 2° relative. Assuming a rare autosomal-dominant mutation, the BF is approximately

$$BF \approx BF_c \times \frac{1}{2} \left(1 + \frac{f_1^{c_1} S_1^{1-c_1}}{f_0^{c_1} S_0^{1-c_1}} \times \left(1 + \frac{f_1^{c_2} S_1^{1-c_2}}{f_0^{c_2} S_0^{1-c_2}} \right) \right), \quad (\text{A.2})$$

assuming π is small relative to all pairs of products of f_1/f_0 and S_1/S_0 . The recursive structure suggests how higher degree relatives are included.



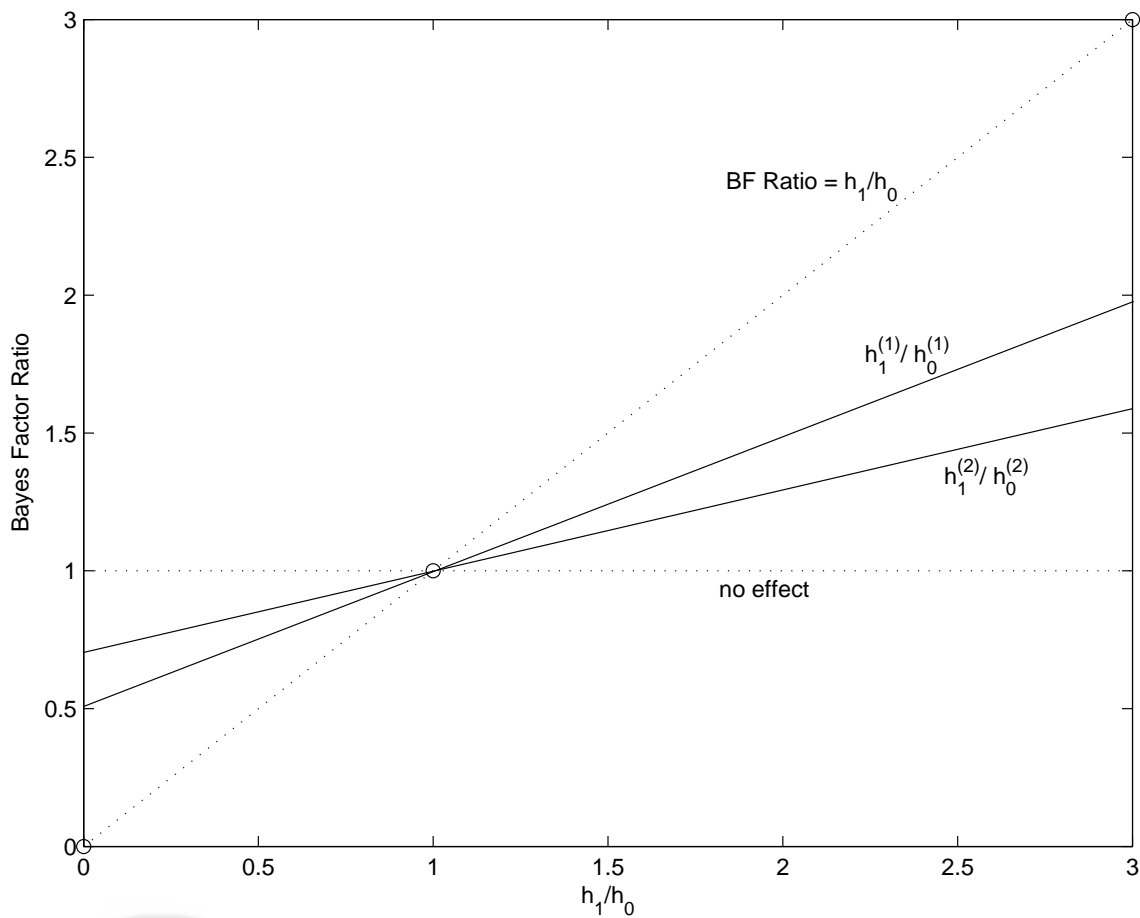


Figure 1. Bayes Factor Ratio in 1° and 2° relatives as a function of the BF ratio in the consultand (which is the hazard ratio)

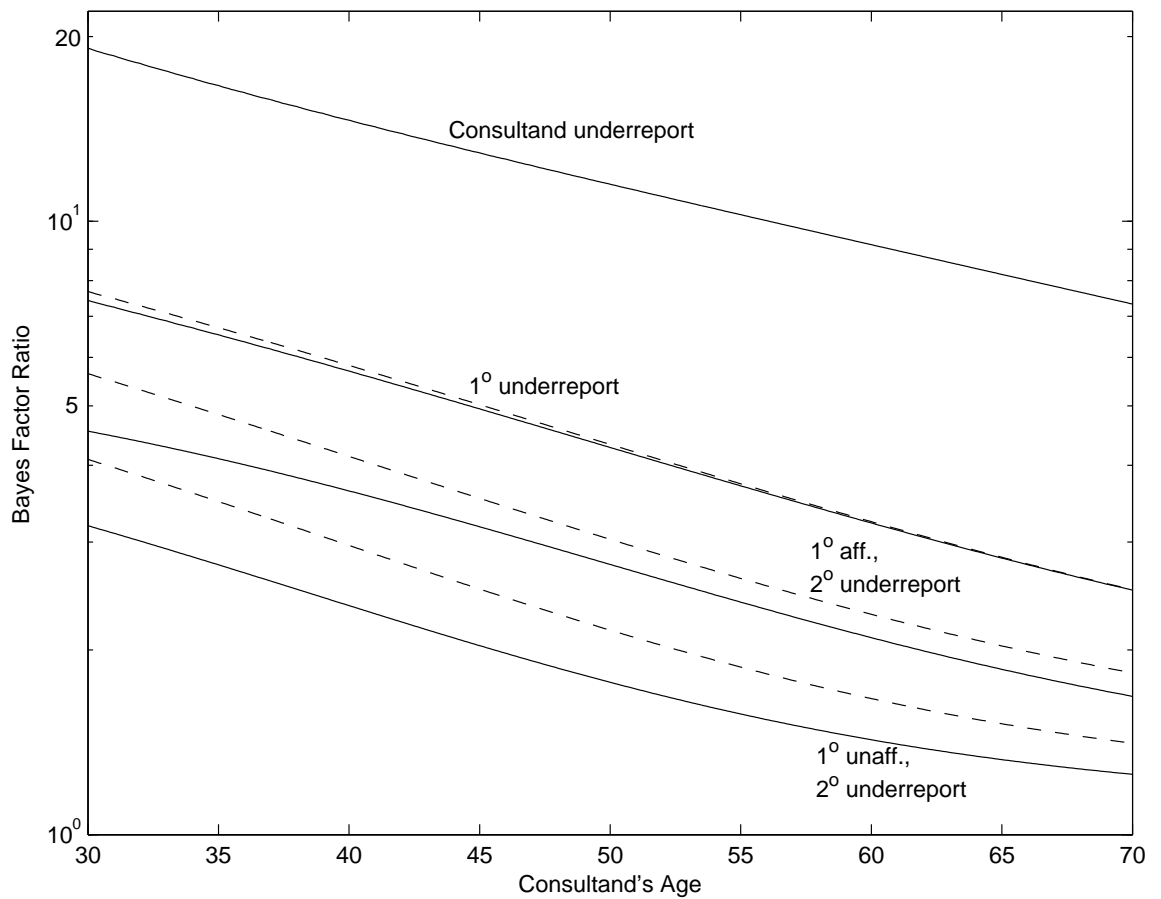


Figure 2. Distortion of the Bayes Factor for Ashkenazi Jewish BRCA1 carrier probability from BRCAPRO by underreporting breast cancer diagnoses in the consultand, 1° , and 2° relatives (for consultand 10,20 years younger than $1^\circ, 2^\circ$ relatives respectively). Solid line is the exact BF ratio, dashed line is the BF ratio based on the rare mutation approximation.

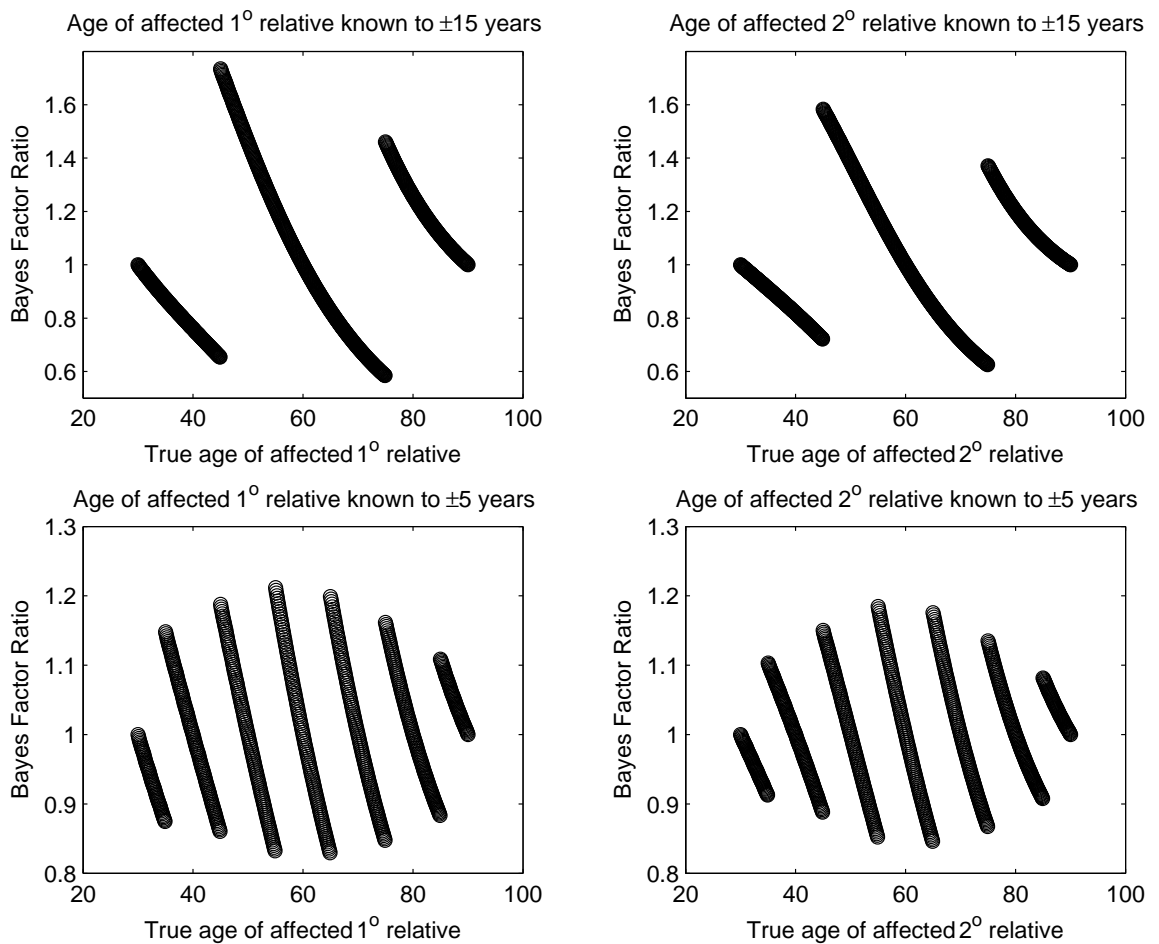


Figure 3. Distortion of the Bayes Factor for Ashkenazi-Jewish BRCA1 carrier probability from BRCAPRO caused by rounding error in age of diagnosis for affected 1^o, 2^o relatives. Rounding is to nearest 30 (± 15 years) or nearest 10 (± 5 years). The 2^o relatives have the 1^o relative known affected at the true age of diagnosis of the 2^o relative.

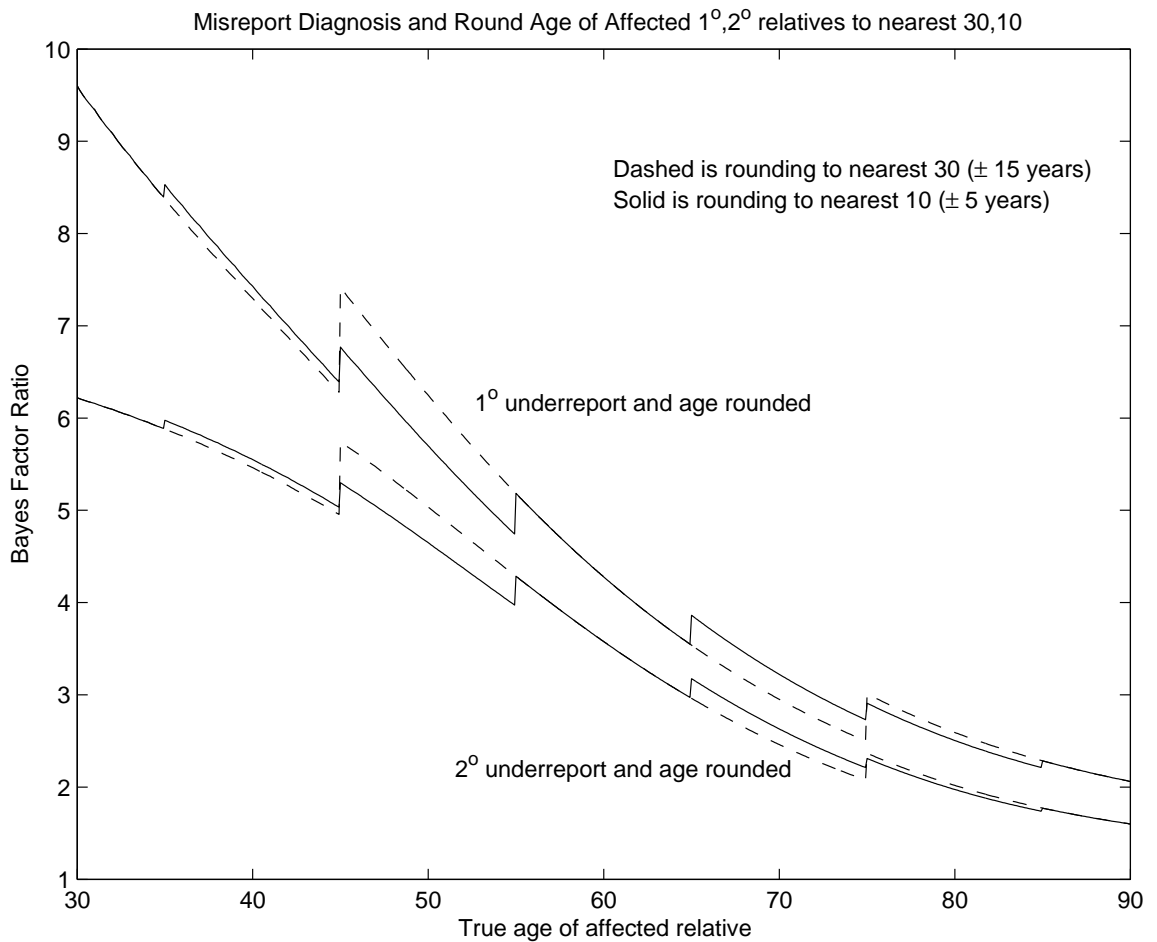


Figure 4. Effect of underreporting diagnosis and rounding error in age of diagnosis for affected 1°, 2° relatives on the Bayes Factor for Ashkenazi-Jewish BRCA1 carrier probability from BRCAPRO. The top two lines are for underreporting the 1° relative, and bottom two for the 2° relative. For the 2° relatives, they have 1° relative known affected at the true age of diagnosis of the 2° relative. Solid line is rounding to nearest 10, dashed line is nearest 30.