Accuracy of MSI testing in predicting germline mutations of MSH2 and MLH1: a case study in Bayesian meta-analysis of diagnostic tests without a gold standard

SINING CHEN*
Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA
schen46@jhmi.edu

PATRICE WATSON
Hereditary Cancer Institute, Creighton University School of Medicine, Omaha, NE, USA

GIOVANNI PARMIGIANI
Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA, and Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

SUMMARY
Microsatellite instability (MSI) testing is a common screening procedure used to identify families that may harbor mutations of a mismatch repair (MMR) gene and therefore may be at high risk for hereditary colorectal cancer. A reliable estimate of sensitivity and specificity of MSI for detecting germline mutations of MMR genes is critical in genetic counseling and colorectal cancer prevention. Several studies published results of both MSI and mutation analysis on the same subjects. In this article we perform a meta-analysis of these studies and obtain estimates that can be directly used in counseling and screening. In particular, we estimate the sensitivity of MSI for detecting mutations of MSH2 and MLH1 to be 0.81 (0.73–0.89). Statistically, challenges arise from the following: (a) traditional mutation analysis methods used in these studies cannot be considered a gold standard for the identification of mutations; (b) studies are heterogeneous in both the design and the populations considered; and (c) studies may include different patterns of missing data resulting from partial testing of the populations sampled. We address these challenges in the context of a Bayesian meta-analytic implementation of the Hui–Walter design, tailored to account for various forms of incomplete data. Posterior inference is handled via a Gibbs sampler.

Keywords: Diagnostic test; Hereditary nonpolyposis colorectal cancer; Microsatellite instability; Sensitivity; Specificity.

*To whom correspondence should be addressed. Current address: Division of Clinical Trials and Biometry, 550 North Broadway, Suite 1103, Baltimore, MD 21205, USA.
1. INTRODUCTION

Hereditary nonpolyposis colorectal cancer (HNPCC) (Lynch and de la Chapelle, 1999) is the most common hereditary cancer syndrome. It can be caused by a deleterious germline mutation of one of the mismatch repair (MMR) genes, mainly \textit{MSH2} and \textit{MLH1}. These mutations confer substantially increased lifetime risk of cancer of the colorectum and other sites (Vasen \textit{et al}., 2001; Lin \textit{et al}., 1998). Microsatellite instability (MSI) is the amplification or deletion within DNA microsatellites (tandemly repeated sequences that appear throughout the genome). This phenotype is caused by a general failure of the DNA MMR system. All HNPCC cases, i.e., cases with inherited defects in the MMR genes, have a strong tendency to develop tumors with this phenotype. For a few years since MSI was discovered in association with HNPCC (Thibodeau \textit{et al}., 1993), this phenomenon was not uniformly characterized. A variety of laboratories developed their own tools for measuring MSI, and related studies indicated that many markers could be used to distinguish MSI tumors from microsatellite stable (MSS) ones. In 1998, ‘The International Workshop on Microsatellite Instability and RER Phenotypes in Cancer Detection and Familial Predisposition’ was held to develop a consensus definition of MSI and to establish unifying criteria for its measurement (Boland \textit{et al}., 1998). As a result, the following five markers are now used for the identification of MSI tumors: \textit{BAT26}, \textit{BAT25}, \textit{D5S346}, \textit{D2S123}, and \textit{D17S250}. Instability in more than two of the five markers indicates an MSI tumor. To date, MSI is found in more than 70% of HNPCC-related cancers (Liu \textit{et al}., 2000) and up to 10–20% of sporadic colorectal cancers (Salovaara \textit{et al}., 2000; Lothe \textit{et al}., 1993). MSI testing is aimed at detecting MMR deficiencies and has become widely used as a prescreening procedure for possible HNPCC germline mutations, thanks to the simpler and less expensive technique compared to most mutation analyses (Boland \textit{et al}., 1998).

A reliable estimate of sensitivity and specificity of MSI for detecting germline mutations of MMR genes is critical in genetic counseling and colorectal cancer prevention. In particular, knowing one’s probability of carrying an inherited deleterious mutation on the \textit{MSH2} and \textit{MLH1} genes prior to genetic testing by mutation analysis is important for decision making about genetic testing, disease prophylaxis, family planning, and more. As many colorectal cancer patients undergo MSI testing before genetic analysis, pretest carrier probabilities need to incorporate information about MSI testing, as done in commonly used prediction software such as BayesMendel (Chen \textit{et al}., 2004). This requires estimates of the sensitivity of MSI, defined as the probability of a subject’s tumor sample being microsatellite unstable (MSI = 1) given he/she is carrying a deleterious germline mutation of \textit{MSH2} or \textit{MLH1} (Mut = 1), and of the specificity of MSI, defined as the probability of the subject’s tumor sample being MSS (MSI = 0) given he/she is not carrying a mutation (Mut = 0). Also, accurate estimates of the test properties along with related costs are determinants in designing screening and surveillance programs for colorectal cancer (Ramsey \textit{et al}., 2001; Reyes \textit{et al}., 2002).

Several investigations have reported testing a set of tumors both for germline mutations and for MSI, and then comparing the results. Among these, studies using tumors from high-risk colorectal cancer cases have relatively small sample sizes. Yet accurate risk prediction and genetic counseling require a reliable estimate of MSI sensitivity and specificity. In this article we perform a meta-analysis of these studies and obtain estimates of sensitivity and specificity as defined above. Statistically, challenges arise from the following: (a) traditional mutation analysis methods used in published studies cannot be considered a gold standard for the identification of mutations (Yan \textit{et al}., 2000); (b) studies are heterogeneous in both the design and the populations considered; and (c) studies may include different patterns of missing data resulting from partial testing of the populations sampled. We address these challenges using a Bayesian meta-analysis motivated by the Hui–Walter approach (Hui and Walter, 1980). This approach exploits the availability of multiple studies that sample populations with different mutation prevalences, to obtain estimates of sensitivity and specificity of two tests without requiring either one to be a gold
standard. We develop specific adaptations designed to accommodate incomplete data and heterogeneity in the experimental designs of the available studies. Posterior inference is handled via a Gibbs sampler.

2. DATA AND METHODS

For our meta-analysis, we considered published studies that involved testing colorectal cancer tumor tissues for MSI and/or germline mutations. We initially identified a number of such studies by interviewing experts, searching on PUBMED, searching reference sets of previously identified papers, and searching citation indices. Not all studies with such results could be included in the analysis. The reason for exclusion and a more detailed description of each study are provided in the Appendix. We then selected studies with clearly reported test results. Examining the results of the literature search, the following issues arose.

The first issue is the absence of a gold standard. In the published studies, tumor samples were tested with a variety of mutation analysis strategies including direct sequencing alone, denaturing gradient gel electrophoresis (DGGE) alone, single-strand conformation analysis (SSCA) alone, polymerase chain reaction (PCR) + SSCA + heteroduplex analysis (HA), DGGE + reverse transcriptase-polymerase chain reaction (RT-PCR) + protein truncation test (PTT), and many others. We discuss the handling of these testing strategies in Section 4. For now, we refer to these tests collectively as 'mutation analysis'. Studies have demonstrated that these conventional mutation testing strategies are not sensitive to certain complex mutations (Wang et al., 2001; Sophie et al., 2001), or large genomic deletions and rearrangements (Yan et al., 2000; Payne et al., 2000; Nakagawa et al., 2002; Wang et al., 2003; Renkonen et al., 2003). The mutations that are likely to be missed by conventional methods make up a significant fraction of all mutations related to colorectal cancers. Because of this limitation, in these studies, the true genotype is not always observed. Using germline testing results as the gold standard can lead to biased estimates of MSI sensitivity and specificity.

The second issue is population heterogeneity both between and within studies. Several studies we included tested families that can be categorized into different groups according to certain clinical criteria. It is likely that the prevalence of deleterious mutation differs from group to group, but the authors often chose not to report the results by groups but rather gave aggregate results. Meanwhile, other studies may focus on a single more homogeneous population, but the populations may be different across studies. Overall, this scenario is a challenge but also a strength because the existence of such strata allows us to account for the lack of a gold standard, as we will discuss later.

The third issue is incompleteness in the data. An ideal study is one where the same set of subjects are tested both for MSI and germline mutations. However, not all studies have complete data in this sense. In Terdiman et al. (2002), Wang et al. (2003), Loukola et al. (1999a), and Salovaara et al. (2000), it was expected that the yield of mutations would be low among MSI-negative tumors and only those samples that tested positive for MSI were genotyped. In Cederquist et al. (2001), only 23 out of the 35 MSI positives were genotyped. Ponz de Leon et al. (2004) only gave mutation analysis results, whereas Percesepe et al. (2001), Salahshor et al. (1999), and Dietmaier et al. (1997) only gave MSI results. While the data in these studies are incomplete, they may provide important information on MSI accuracy. Therefore, the inclusion of such data was judged to be worth the additional modeling effort required.

In our analysis, we stratified the study populations into a high-risk group and a low-risk group. The high-risk group includes families that fulfilled the Amsterdam criteria, or met any modified clinical guidelines for the purpose of identifying HNPCC families. In addition, we have included in the high-risk group all families and individual cases that can be classified as being at significantly elevated risk, based on descriptors, such as ‘with age of onset before 35’, ‘with two or more close relatives with colorectal cancer’, or ‘exhibiting familial clustering of colorectal cancer (strong and positive family history)’. The low-risk group includes cases from population-based registries. The grouping we use here attempts, within the limits imposed by the information available, to distinguish between data gathered in high-risk clinical
settings, most of which is self-referred, and data gathered using population-based strategies. While our approach would have been applicable, and potentially more efficient, to a larger set of ordered categories of risk, we stopped at two because several studies tested a possibly heterogeneous sample and only reported aggregate results, and to assure that each group has a reasonable sample size. The stratified data are shown in Table 1.

If all missing counts in Table 1 were observed, then the stratified data would fit into the framework of a Hui–Walter design (Hui and Walter, 1980). In this design, the sensitivity and specificity of MSI testing can be estimated without ascertaining the true genetic status of the subjects, by exploiting the fact that the strata that are known to have different prevalences. Data consist of a three-way tabulation of the results of two tests and the stratum indicator. The Hui–Walter approach relies on two key modeling assumptions: independence of the testing modalities and different prevalences among the high- and low-risk groups (Parmigiani, 2002; Johnson et al., 2001). Further discussion of these assumptions is in Section 4.

To write the likelihood function, we define the multinomial probabilities of an individual falling into the four cells MSI = 1 & Mut = 1, MSI = 1 & Mut = 0, MSI = 0 & Mut = 1, and MSI = 0 &

Table 1. A list of the studies included. Data are stratified into a high-risk group in the top portion and a low-risk group in the bottom portion. In each row, data are summarized in four columns by MSI test results (MSI = 0 or 1) and mutation analysis results (Mut = 0 or 1). Missing counts due to samples without one of the two tests are represented by x and y. For example, x3 stands for the potential number of mutation analysis positive subjects in Ponz de Leon et al. (2004) who would have been tested positive for MSI. Among the MSI-positive group of 35 in Cederquist et al. (2001), mutation analysis was performed on 23 samples among which eight were found to harbor germline mutations. No mutation analysis was performed on the other 12 samples. x1 represents the potential number of mutation positives among these 12 untested subjects.
Mut = 0 as \( a, b, c, \) and \( d, \) respectively. Subscripts ‘h’ and ‘l’ will denote the high-risk group and low-risk group, respectively, while subscript \( + \) and \( - \) will denote true carrier status. For example, \( a_{h+} \) are MSI positive, mutation analysis positive, and true carriers in the high-risk group. We denote the sensitivity and specificity of MSI testing by \( \beta_M \) and \( \alpha_M, \) those of the mutation analysis by \( \beta_T \) and \( \alpha_T, \) and the mutation prevalences in the high-risk and low-risk groups by \( \pi_h \) and \( \pi_l. \) Then, using the independence assumption for the two testing modalities, we can write for the high-risk group,

\[
\begin{align*}
    a_h & \equiv a_{h+} + a_{h-} = \pi_h \beta_M \beta_T + (1 - \pi_h)(1 - \alpha_M)(1 - \alpha_T), \\
    b_h & \equiv b_{h+} + b_{h-} = \pi_h \beta_M (1 - \beta_T) + (1 - \pi_h)(1 - \alpha_M)\alpha_T, \\
    c_h & \equiv c_{h+} + c_{h-} = \pi_h (1 - \beta_M) \beta_T + (1 - \pi_h)\alpha_M (1 - \alpha_T), \\
    d_h & \equiv d_{h+} + d_{h-} = \pi_h (1 - \beta_M)(1 - \beta_T) + (1 - \pi_h)\alpha_M \alpha_T.
\end{align*}
\]

Similar relationships apply to the low-risk group. We also use \( \theta \) to denote the parameter vector \( (\alpha_T, \alpha_M, \beta_T, \beta_M, \pi_h, \pi_l) \) and \( O \) to denote the totality of observed variables in Table 1.

Using this notation, we could write the contribution from each study to the likelihood function, depending on the stratum and missing data pattern. For example, the contribution from Bapat \textit{et al.} (1999) was

\[
a_h^{16} b_h^{1.2} c_h^{20} d_h^{2},
\]

that from Terdiman \textit{et al.} (2001) is

\[
a_h^{21} b_h^{11} (c_h + d_h)^{63},
\]

and that from Percesepe \textit{et al.} (2001) is

\[(a_l + b_l)^{28} (c_l + d_l)^{308}.
\]

Assuming independence of study results conditional on \( \theta, \) the likelihood is the product of the contribution from each study, that is,

\[
L(\theta|O) = \{ \underbrace{a_h^{79} b_h^{63} c_h^{10} d_h^{134}}_{\text{Bapat}} \} \{ \underbrace{a_h^{8} b_h^{15} (a_h + b_h)^{12} (c_h + d_h)^{45}}_{\text{Terdiman}} \} \{ \underbrace{(a_h + c_h)^{89} (b_h + d_h)^{75}}_{\text{Percesepe}} \} \times \{ \underbrace{(a_l + b_l)^28 (c_l + d_l)^{20}}_{\text{Bapat}} \} \{ \underbrace{(a_l + b_l)^28 (c_l + d_l)^{20}}_{\text{Terdiman}} \} \{ \underbrace{(a_l + b_l)^{22} (c_l + d_l)^{159}}_{\text{Percesepe}} \} \times \{ \underbrace{(a_l + b_l)^{18} (c_l + d_l)^{130}}_{\text{Bapat}} \} \{ \underbrace{(a_l + b_l)^{48} (c_l + d_l)^{149}}_{\text{Terdiman}} \} \{ \underbrace{(a_l + b_l)^{18} (c_l + d_l)^{130}}_{\text{Percesepe}} \} \{ \underbrace{(a_l + b_l)^{48} (c_l + d_l)^{149}}_{\text{Terdiman}} \} \times \{ a_l^{106} b_l^{45} (a_l + b_l)^{68} (c_l + d_l)^{1512}, \}
\]

where each bracketed expression corresponds to a study. Under our modeling assumptions, all six parameters \( \beta_M, \alpha_M, \beta_T, \alpha_T, \pi_h, \) and \( \pi_l \) are identified, as long as the constraint \( \pi_h > \pi_l \) holds.

We drew inferences about \( \theta \) via a Bayesian analysis. Advantages include the possibility to derive exact probability intervals without relying on asymptotic approximations, and ease of handling both the missing data patterns and the parametric constraint required for identifiability. Within this constraint, we adopted a uniform prior on all components of \( \theta, \) with the exception of the specificity \( \alpha_T \) of mutation analysis, which we assumed to be 100%, a value likely to be close to the true specificity. However, as a sensitivity analysis, we also reported results using a uniform prior on \( \alpha_T, \) the opposite extreme in term of informativeness.
Table 2. Notation: the number in each cell can be broken down into those with a true germline mutation (subscript ‘+’) and those without (subscript ‘−’). High risk and low risk are indicated by the letter ‘h’ and ‘l’, respectively, in the subscript. The superscripts in parentheses signify contributions from studies with different missing data structures.

<table>
<thead>
<tr>
<th>Mut = 1</th>
<th>Mut = 0</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High risk</strong></td>
<td></td>
</tr>
<tr>
<td>MSI = 1</td>
<td>(n_{a,h+} = n_{a,h+}^{(1)} + n_{a,h+}^{(2)} + n_{a,h+}^{(3)})</td>
</tr>
<tr>
<td>MSI = 0</td>
<td>(n_{c,l+} = n_{c,l+}^{(1)} + n_{c,l+}^{(2)})</td>
</tr>
</tbody>
</table>

| **Low risk** |                     |                     |
| MSI = 1  | \(n_{a,l+} = n_{a,l+}^{(1)} + n_{a,l+}^{(2)}\) | \(n_{b,l+} = n_{b,l+}^{(1)} + n_{b,l+}^{(2)}\) |
| MSI = 0  | \(n_{c,l+} = n_{c,l+} \) | \(n_{d,l+} = n_{d,l+}^{(1)} + n_{d,l+}^{(2)}\) |

From Bayes theorem, the posterior is proportional to expression (2.1) wherever the prior is nonzero. Computation was simplified by introducing two sets of auxiliary variables: the counts of test results for subjects who have not been tested by both tests, and the counts of true genotypes for all subjects. Denoting auxiliary variables by \(A\), we implemented a Markov chain Monte Carlo algorithm to draw from the distribution of \((A, \theta)\) given \(O\) (Tanner and Wong, 1987).

Our algorithm requires an expression for the so-called complete data likelihood, that is, \(p(A, O|\theta)\). This is given by

\[
a_{h+}^{n_{a,h+}} \cdot a_{h-}^{n_{a,h-}} \cdot b_{h+}^{n_{b,h+}} \cdot b_{h-}^{n_{b,h-}} \cdot c_{h+}^{n_{c,h+}} \cdot c_{h-}^{n_{c,h-}} \cdot d_{h+}^{n_{d,h+}} \cdot d_{h-}^{n_{d,h-}} \cdot a_{l+}^{n_{a,l+}} \cdot a_{l-}^{n_{a,l-}} \cdot b_{l+}^{n_{b,l+}} \cdot b_{l-}^{n_{b,l-}} \cdot c_{l+}^{n_{c,l+}} \cdot c_{l-}^{n_{c,l-}} \cdot d_{l+}^{n_{d,l+}} \cdot d_{l-}^{n_{d,l-}}.
\]

(2.2)

Each exponent in (2.2) arises as the sum of contributions from studies with different missing data structures, which needed to be simulated separately, requiring additional auxiliary variables defined in Table 2. To explain the significance of the auxiliary variables, for example, we observe in the high-risk group that 200 subjects are MSI positive and mutation analysis positive (MSI+ & Mut+), and 177 are MSI+ & Mut−. This corresponds to the first two factors in the expression (2.1), \(a_{h+}^{200} b_{h-}^{177}\). After we assume the carrier status known, among the 200 MSI+ & Mut+ subjects, \(n_{a,h+}^{(1)}\) would be the true carriers and \(n_{a,h-}^{(1)} = 200 - n_{a,h+}^{(1)}\) the true noncarriers. Similarly, among the 177 MSI+ & Mut− subjects, \(n_{b,h+}^{(1)}\) would be true carriers and \(n_{b,h-}^{(1)}\) noncarriers. Thus, the corresponding factors in the complete data likelihood were written as \(a_{h+}^{n_{a,h+}^{(1)}} a_{h-}^{n_{a,h-}^{(1)}} b_{h+}^{n_{b,h+}^{(1)}} b_{h-}^{n_{b,h-}^{(1)}}\) with the constraints \(n_{a,h+}^{(1)} + n_{a,h-}^{(1)} = 200\) and \(n_{b,h+}^{(1)} + n_{b,h-}^{(1)} = 177\). In other words, \(a_{h+}^{200} b_{h-}^{177}\) is the result of integrating out the auxiliary variables \(n_{a,h+}^{(1)}, n_{a,h-}^{(1)}, n_{b,h+}^{(1)}, n_{b,h-}^{(1)}\) from \(a_{h+}^{n_{a,h+}^{(1)}} a_{h-}^{n_{a,h-}^{(1)}} b_{h+}^{n_{b,h+}^{(1)}} b_{h-}^{n_{b,h-}^{(1)}}\) within the constraints. Details of the Gibbs sampler and the various constraints placed on the components are available from the first author.
3. Results

Posterior means and 95% credible intervals (CIs) on the parameters of interest are shown in Table 3. The estimated specificity of MSI testing was 92% (95% CI, 90–94%), indicating that MSI is a more specific test than previously suggested (Salovaara et al., 2000; Lothe et al., 1993). Combined with an estimated sensitivity of 81%, our estimates translate into a positive predictive value (PPV) of 0.93 (95% CI, 0.90–0.95) among high-risk families. However, the PPV of MSI is drastically lower among unselected colorectal cancer cases: we estimate it to be 0.34 (95% CI, 0.24–0.48). Both PPV estimates are consistent with practical clinical experiences in MSI screening. Our prevalence estimate of true carriers among unselected colorectal cancer, 3–7%, coincides with previous estimates (Aaltonen et al., 1994, 1998).

We estimated the sensitivity of mutation analysis to be 62%. While this is a modest value, our estimates are comparable to those reported for similar mutation analysis techniques for other large cancer susceptibility genes. For example, Ford et al. (1998) reported an estimated sensitivity of 63% (51–77%) for direct sequencing in detecting mutations on BRCA1. They also find that the sensitivity for other mutation screening techniques is not significantly different from that of direct sequencing. While clinical use of sequencing-based testing is still justified in view of the much higher specificity of sequencing compared to MSI testing, our results further emphasize the need for cautious interpretation of negative test results in families with a large number of colorectal cancer cases (Yan et al., 2000).

At each iteration of the Gibbs sampler, we generated a draw of all entries in Table 2. These draws can be used to make inferences on the missing counts in Table 1. For example, in the Ponz de Leon et al. (2004) study, $x_3$ was imputed as 72 on average by the Gibbs sampler, which means that 72 out of the 89 mutation-positive subjects in the study would have been tested positive for MSI. In addition, $x_1$ was on average 7 and $x_4$ was on average 24. The other missing counts could be estimated up to a sum. The quantities $x_2 + x_5 + x_6$, $y_2 + y_4 + y_6$, and $y_1 + y_3 + y_5 + y_7 + y_8$ had posterior means around 38, 14, and 10, respectively.

With the Gibbs sampler, in addition to inferences on missing portions of the original studies, we also obtained inferences on the true genotypes for all subjects. Computing the total of each of the rightmost four columns of Table 1, and then breaking them down using imputed carrier status results, we produced Table 4. From Table 4 we can see that although MSI is quite specific, because of the low prevalence of mutations in the low-risk population, there will still be a significant portion of false positives (an estimated 133 out of 202).

As a sensitivity analysis, we relaxed the assumption that the mutation analysis specificity is 100%, that is, that $\alpha_T \equiv 1$, in favor of a uniform prior. Results are also shown in Tables 3 and 4. The posterior mean of the mutation analysis specificity was 0.97 (95% CI, 0.91–1.0). Comparing the rest of the parameter estimates with those obtained under $\alpha_T \equiv 1$, we can see that this departure from the assumption had

| Table 3. Summary of inferences on parameters. The left part of the table is obtained by setting $\alpha_T$ to 1. The right part is obtained by placing a uniform prior on $\alpha_T$. |
|---------------------------------|---------------------------------|
|                                 | $\alpha_T = 1$                  | $\alpha_T \sim U(0, 1)$         |
|                                 | Mean (95% PI)                    | Mean (95% PI)                    |
| MSI specificity $\alpha_M$     | 0.92 (0.90, 0.94)                | 0.92 (0.90, 0.93)                |
| MSI sensitivity $\beta_M$      | 0.81 (0.73, 0.89)                | 0.85 (0.74, 0.96)                |
| Mutation analysis specificity $\alpha_T$ | 1                               | 0.97 (0.91, 1.0)                |
| Mutation analysis sensitivity $\beta_T$ | 0.62 (0.56, 0.67)               | 0.62 (0.57, 0.68)               |
| Prevalence in high-risk group $\pi_h$ | 0.56 (0.50, 0.63)               | 0.53 (0.46, 0.61)               |
| Prevalence in low-risk group $\pi_l$ | 0.049 (0.032, 0.071)             | 0.042 (0.022, 0.064)             |
Table 4. The data table broken down by imputed genotype. The top table is imputed under the constraint that the specificity of mutation analysis is 1, the bottom without the constraint. In each cell, the first number is the expected number of true mutation carriers of all subjects in that corresponding column in Table 1; the second number is the number of noncarriers.

<table>
<thead>
<tr>
<th></th>
<th>High risk</th>
<th>Low risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mut = 1</td>
<td>Mut = 0</td>
</tr>
<tr>
<td>$a_T = 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI = 1</td>
<td>279 + 0</td>
<td>171 + 35</td>
</tr>
<tr>
<td>MSI = 0</td>
<td>65 + 0</td>
<td>41 + 400</td>
</tr>
<tr>
<td>$a_T &lt; 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI = 1</td>
<td>278 + 1</td>
<td>168 + 38</td>
</tr>
<tr>
<td>MSI = 0</td>
<td>51 + 13</td>
<td>32 + 410</td>
</tr>
</tbody>
</table>

a small impact on the posterior inference. The CIs were only slightly wider, as a result of increased uncertainty. This suggests that the likelihood is informative about the parameters of interest. Also, since some of the true positives (in terms of mutation analysis) under the specificity assumption were imputed as false positives, the posterior distributions of the proportions of carriers in the two groups shifted slightly to the left. The MSI sensitivity estimate increased slightly from 0.81 (0.73–0.89) to 0.85 (0.74–0.96). Comparing the upper portion of Table 4 to the lower portion, in all the mutation-positive cells, a few were imputed as false positives in the lower portion. However, such imputation occurred more often among MSI-negative subjects than among MSI-positive ones, causing a small increase in the overall MSI sensitivity. The only sensitive figure was the posterior mean of $y_1 + y_3 + y_5 + y_7 + y_8$, which increased from 12 to 47 when a uniform prior was assigned to $a_T$.

To further explore the sensitivity of our inferences to the choice of priors, we considered the study by Aaltonen et al. (1994), who found 25 out of 29 HNPCC patients to be MSI+. While their definition of an ‘HNPCC patient’ is different from the gold standard used in our paper, it can still provide a plausible reference for a sensitivity analysis. Thus, we explored using a Beta(25, 4) as the prior on $\beta_M$. The resulting posterior mean and 95% probability interval shifted from 0.81 (0.73, 0.89) to 0.85 (0.76, 0.93), suggesting robustness of our findings in terms of both location and spread. This suggests that the width of the probability intervals we report should adequately convey the level of uncertainty about the parameters based on currently available evidence.

4. DISCUSSION

In this paper we presented a meta-analysis of the accuracy of MSI testing in detecting $MSH2$ and $MLH1$ mutations. Our meta-analysis overcame the difficulty posed by the fact that no gold standard is available in the published literature about this issue, and provided estimates that can be directly used in genetic counseling and colorectal cancer screening. Our methodology was a Bayesian meta-analytic implementation of the Hui–Walter approach, adapted to account for various forms of incomplete data.

We found MSI to be a very sensitive and specific indicator of germline mutations. The estimates of sensitivity and specificity of MSI in detecting $MSH2$ and $MLH1$ mutations will be incorporated in the latest version of the carrier probability package BayesMendel (Chen et al., 2004) and the risk prediction model CRCAPRO. The false-positive fraction associated with MSI testing can be large among unselected colorectal cancers. We estimated the sensitivity of mutation analysis at a modest 62%. This result suggests using caution when interpreting negative genetic testing results for potential mutation carriers.
One of the limitations of our study is that we only considered the \textit{MSH2} and \textit{MLH1} genes. Other, less prevalent HNPCC genes have been found. For example, \textit{MSH6} has a mutation prevalence of about 1.7% among unselected colorectal or endometrial cancer cases (Goodfellow et al., 2003). At the present time, it would be difficult to consider other genes because data from published studies involving such genes are scarce.

A further concern is that the published studies have used a variety of mutation analysis techniques. We pooled all these as though they had the same sensitivity. The sensitivity of a particular mutation search method depends on the true spectrum of mutations, particularly the relative frequency of each type of mutation. In the context of HNPCC mutations, the pathogenic mutations discovered to date in \textit{MLH1} and \textit{MSH2} are varied in type, and the true spectrum of mutations is still unknown. Systematic studies aimed at comparing sensitivity among the mutation search methods are not presently available, and thus there is no evidence as to which methods are more sensitive over the true mutation spectrum than others. In addition, there are practical reasons why allowing for different sensitivities is unlikely to be successful in our meta-analysis: mutation screening strategies differed across studies, and in many cases were not reported in sufficient detail to allow stratified analysis. Also, if we were to assign each of the testing strategies a separate sensitivity and specificity, the sample sizes for each strategy would become too small.

To further explore this issue, we designed a simple simulation experiment where subjects are tested with genotyping approaches with different sensitivities. Our results indicated that, as long as the distributions of genotyping methods in the high-risk and low-risk populations are similar, estimates of MSI specificity and sensitivity remain unbiased. Specifically, we simulated a data set where we tested a high-risk population of 800 (prevalence, 0.6) and a low-risk population of 1700 (prevalence, 0.05) with a pre-screen with sensitivity of 0.78 and specificity of 0.92. In each population, we randomly chose half of the subjects to be subsequently genotyped with sensitivity 0.60 and the other half with sensitivity 0.80 (both specificities are 1). To imitate the missing data situation in our meta-analysis, those who were prescreened negative did not receive a second test. With 100 simulations, all parameter estimates were unbiased and the mean relative errors were all of the order of magnitude $10^{-3}$ or less.

A key assumption in our analysis is that the results of MSI testing and mutation analysis are independent given the true genotype. In other words, we assume that if a specific mutation is difficult to detect by mutation analysis, the same mutation is not necessarily difficult to detect by MSI testing. As we write, there is no strong biological evidence that such dependence may arise. The errors associated with MSI with respect to the carrier status of \textit{MSH2} and \textit{MLH1} are likely to arise from links between MSI and changes in the MMR system other than a mutation on \textit{MSH2} or \textit{MLH1}. On the other hand, limitations of sequencing and other forms of conventional mutation analysis arise from the complex spectrum of mutations on \textit{MSH2} and \textit{MLH1} themselves. Thus, the unknown mechanisms driving test errors are likely to be different, though further validation of this assumption would be helpful.

A second key assumption in our analysis is that the prevalence of mutations in the two groups should be different. This is not a restrictive assumption, given the way the groups were defined. However, groups were separated based on a sensible but subjective choice and it is conceivable that the results may be sensitive to this choice. To address this issue, we performed a simulation experiment. The results indicated that different groupings of the data according to risk do not change the unbiasedness of the estimates of MSI sensitivity and specificity, as long as the risk in one group is higher than the other. In this experiment, in addition to a homogeneous high-risk group of prevalence 0.6 and a low-risk group of prevalence 0.05, we simulated a group of ‘medium-high risk’, with prevalence of 0.3. The rest of the setup was similar to the simulation described above. In case 1 we pooled the medium-high-risk group with the high-risk group; in case 2 we pooled it with the low-risk group. In both cases, the posterior mean was unbiased for all parameters.
ACKNOWLEDGMENTS

The work by Sining Chen and Giovanni Parmigiani is supported by grants NCI-P50CA62924, NCI R01CA105090, and NCI-P30CA06973. The work by Patrice Watson is supported by NIH Grant #1U01 CA86389. The authors thank Hormuzd A. Kakti, Kenneth Kinzler, and Jiang-wen Zhang for helpful comments.

APPENDIX

Summary of studies

While we included as many studies as possible, we had to exclude studies where the subjects are a mixture of high-risk and low-risk cases with no breakdown given, such as Ruszkiewicz et al. (2002) and Banno et al. (2003). We disregarded all immunohistochemistry (IHC) results where reported because the literature suggests a very high correlation between MSI and IHC (Lindor et al., 2002; Debniak et al., 2000; Marcus et al., 1999), which may violate the conditional independence assumption.

In the searches for germline mutations, it is common for studies to sequence those samples that showed alterations that first underwent PTT or DGGE as a confirmation step. However, to our knowledge there was no gain of sensitivity or specificity in this extra step. A few studies used novel methods to search for germline mutations that are hard to detect by conventional methods (Renkonen et al., 2003; Nakagawa et al., 2003; Wang et al., 2003). Unfortunately, since these technologies are very different from conventional methods, pooling them with the conventional methods may introduce bias. We decided to exclude such studies or to exclude mutations found by using such technologies. However, such data can serve as independent validation of the estimated low sensitivity for conventional methods.

Table A1 lists the reasons for exclusion, while Table A2 gives the characteristics of the 19 studies included in the meta-analysis.

Table A1. Relevant studies and their reasons for being excluded from the meta-analysis

<table>
<thead>
<tr>
<th>Paper</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banno et al. (2003)</td>
<td>Endometrial tumors</td>
</tr>
<tr>
<td>Debniak et al. (2001)</td>
<td>Use of Alu-PCR (no conventional mutation analysis)</td>
</tr>
<tr>
<td>Loukola et al. (1999b)</td>
<td>CR adenomas</td>
</tr>
<tr>
<td>Nakagawa et al. (2003)</td>
<td>Use of multiplex ligation-dependent probe amplification method</td>
</tr>
<tr>
<td>Renkonen et al. (2003)</td>
<td>Use of the SNuPE technique</td>
</tr>
<tr>
<td>Terdiman et al. (2002)</td>
<td>Insufficient presentation of the data</td>
</tr>
</tbody>
</table>

Table A2. Characteristics of studies that were included in the meta-analysis. Note that the Debniak et al. (2000) study and the Dieumegard et al. (2000) study were each subdivided into a high-risk set and a low-risk set, and thus appeared in Table 1 twice

<table>
<thead>
<tr>
<th>Paper</th>
<th>Ethnic</th>
<th>Families</th>
<th>Mutation testing strategy</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bapat et al. (1999)</td>
<td>Mixed</td>
<td>14 AC(^a) + 19 MC(^b)</td>
<td>RT-PCR + PTT</td>
<td></td>
</tr>
<tr>
<td>Calistri et al. (2000)</td>
<td>Italian</td>
<td>13 AC + 7 MG(^c) + 4 fh(^d)</td>
<td>Exon-by-exon SSCP</td>
<td></td>
</tr>
</tbody>
</table>

continued
<table>
<thead>
<tr>
<th>Paper</th>
<th>Ethnic</th>
<th>Families</th>
<th>Mutation testing strategy</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cederquist et al. (2001)</td>
<td>Swedish</td>
<td>Double primary tumor at colorectum or endometrium</td>
<td>DHPLC confirmed by sequencing</td>
<td>MSH6 results excluded</td>
</tr>
<tr>
<td>Debniak et al. (2000)</td>
<td>Polish</td>
<td>11 def.hnppce(^c) (or highly likely) + 15 sporadic late-onset CRC</td>
<td>Sequencing</td>
<td>IHC results discarded</td>
</tr>
<tr>
<td>Dietmaier et al. (1997)</td>
<td>German</td>
<td>148 sporadic CRC</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Dieumegard et al. (2000)</td>
<td>French</td>
<td>15 ICG(^f) + 1 fh + 7 sporadic&lt; 50</td>
<td>SSCP</td>
<td>IHC results discarded</td>
</tr>
<tr>
<td>Lamberti et al. (1999)</td>
<td>German</td>
<td>47 AC + 5 looser AC</td>
<td>PCR + SCA + HA, etc.</td>
<td>Some data excluded b/c of poor presentation</td>
</tr>
<tr>
<td>Liu et al. (2000)</td>
<td>Swedish</td>
<td>‘Highly suspicious for HNPCC’</td>
<td>DGGE + RT-PCR + PTT</td>
<td>Some AC, some fh, some early CRC, no break-down given</td>
</tr>
<tr>
<td>Loukola et al. (1999)</td>
<td>Finnish</td>
<td>CRC registry</td>
<td>DGGE or Sequencing on MSI+</td>
<td></td>
</tr>
<tr>
<td>Percesepe et al. (2001)</td>
<td>Italian</td>
<td>CRC registry</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Ponz de Leon et al. (2004)</td>
<td>Italian</td>
<td>‘Consistent with HNPCC’</td>
<td>SSCP or Sequencing</td>
<td>MSI results n.a.</td>
</tr>
<tr>
<td>Salahshor et al. (1999)</td>
<td>Swedish</td>
<td>Sporadic CRC</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Salovaara et al. (2000)</td>
<td>Finnish</td>
<td>Sporadic CRC</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Scartozzi et al. (2002)</td>
<td>Italian</td>
<td>At least one of the BG(^g)</td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td>Terdiman et al. (2001)</td>
<td>US</td>
<td>23% met AC, 70% BG</td>
<td>DGGE</td>
<td></td>
</tr>
<tr>
<td>Wahlberg et al. (2002)</td>
<td>US</td>
<td>38 AC or BG + 5 MG + 5 HNPCC-like</td>
<td>Sequencing</td>
<td></td>
</tr>
<tr>
<td>Wang et al. (2003)</td>
<td>German</td>
<td>BG</td>
<td>SSCP, HA, DHPLC followed by sequencing</td>
<td>Semiquantitative multiplex PCR result excluded</td>
</tr>
</tbody>
</table>

CRC = colorectal cancer; SSCP = single-strand conformation polymorphism; DHPLC = denaturing high-performance liquid chromatography; ASO = allele-specific oligonucleotide.
\(^a\)Amsterdam criteria.
\(^b\)Mt. Sinai Hospital criteria.
\(^c\)Modified guidelines.
\(^d\)Positive family history.
\(^e\)Definitive HNPCC.
\(^f\)International Collaborative Group criteria.
\(^g\)Bethesda guidelines.

REFERENCES

Accuracy of MSI testing in predicting germline mutations of MSH2 and MLH1


[Received June 25, 2004; first revision October 14, 2004; second revision January 11, 2005; accepted for publication February 7, 2005]