An Analysis Procedure Illustrated on a Triple Linkage of Use for Prenatal Diagnosis of Myotonic Dystrophy

J. H. RENWICK and D. R. BOLLING

From the London School of Hygiene and Tropical Medicine, Keppel St, London WC1E 7HT, UK and Moore Clinic, Johns Hopkins Hospital Medical School, Baltimore, Md. 21205, USA

Diagnosis of a number of simple genetic conditions in man is now feasible prenatally (Nadler, 1971 and others) hence, for the first time, parents sometimes have the opportunity of requesting a selective abortion. For those simple genetic conditions that cannot be so diagnosed by direct means, indirect diagnosis through the prenatal phenotype for a marker locus known to be linked to the disease locus (Edwards, 1956; Renwick, 1969), is a useful substitute with potentially wide applicability. A special case is the use, in certain families, of the chromosomal sex of the embryo to give a 50% or 0% prognosis of haemophilia for male and female respectively. Attention is now drawn to an autosomal linkage between the myotonic dystrophy locus (Dm) and the ABH secretor locus (Se) that can be used in favourable circumstances to predict fairly accurately the outcome of a particular pregnancy in a family with this dystrophy.

A method of three-locus analysis was required in this work because an additional locus, Lu (for the Lutheran blood group), was involved. Much of the present report concerns an extension—very suitable for this purpose—of a standard Bayesian two-locus analysis (Renwick, 1969, based on Smith, 1959).

Data

On sib-pair analysis of linkage data on the pedigrees of Thomasen (1948), Mohr (1954) found indications of linkage of Dm with Se as scored from the Lewis red-cell phenotype. He found a similar indication with the Lutheran blood group locus (Lu), which, in turn, he knew to be linked to Se (Mohr, 1954; Cook, 1965).

A comprehensive, computer-based analysis has now clarified the real strength of the evidence and has encouraged collection of a further sample. The confirmation afforded by this is the subject of a separate report (Renwick et al., 1971), which also deals with certain problems of scoring the dystrophy and ABH secretion genotypes.

Genetic Model Underlying the Analysis

Some of the assumptions and approximations underlying linkage analysis in man are discussed by Renwick (1971). Only a few of these points require mention here.

1. Intervals between loci are customarily measured as map-lengths because these are additive, whereas recombination fractions cannot be combined in a simple fashion over neighbouring intervals, unless these are short.

2. A morgan is the map-length of any segment of a recovered chromosome strand that has experienced, on average, one crossover event in one meiosis.

3. Chromatid interference is believed to be at a high level in man, leading to the mapping function, 4a = tan⁻¹ 2θ + tanh⁻¹ 2θ, where w morgans is the map-length of an interval that gives a recombination fraction, θ.

4. In women, the susceptibility to crossing-over is, for at least one and perhaps for most autosomal regions, higher than in men (Renwick, 1968). If consistency is provisionally assumed, the current estimate of the ratio, female: male, is 1:40 (95% probability limits 1:04 and 1:98) from unpublished calculations of J. H. Renwick and D. R. Bolling on data from many laboratories.

5. The total autosomal map-length, T, based on chiasma counts will be taken to be 271 morgans in men. (This is an overestimate, hence it errs slightly on the side of safety in reducing the posterior odds on syntenic hypotheses.) The susceptibility-ratio estimate, 1:4, implies a map-length of about 38½ morgans in women.
(6) Meiotic studies in men (M. Hultén, personal communication; P. L. Pearson, personal communication) show more chiasmata on long autosomes than on short ones. The approximation is here made that the relationship is a linear one. The physical lengths given by the Chicago Conference (1966) then offer a basis for partitioning the total autosomal map-length among the autosomal pairs of chromosomes. The resulting estimate of the map-length of the ith chromosome pair will be denoted by \( A_i \) and will be used in calculating prior probabilities as if it were the true value.

**Bayesian Analysis for Three Loci**

An analysis appropriate for multiple loci has been presented by Bolling (1970). Only the three-locus analysis will be given here but the same computer program embraces also the four-locus analysis. A brief outline of the Bayesian framework will be followed by a description of the procedure in more detail.

**Prior Probabilities.** The prior probability of three autosomal loci being on the same chromosome (3-synteny) is \( \sum(A_i/T)^3 \) or about 3/900. The prior probability of 2-on, 1-off (2-synteny; 1-asyn-tency) is \( 3\sum(A_i/T)^3(T-A_i)/T \) or \( 3\sum(A_i/T)^3-3\sum(A_i/T)^3 \). (This happens to be three times the difference between the probabilities for 2-synteny and for 3-synteny.) Its numerical value is about 138/900. In Tables I and II this is partitioned equally between three options, in each of which a different locus is left out of the syntenic group. The probability for 3-synteny is likewise partitioned equally between the three possible orderings of the loci (ignoring orientation throughout). The remaining hypothesis (asyn-tency), under which all three loci are on different autosomes, carries the remaining probability, 759/900.

**Contributions of the Data.** The theorem of Bayes (1763) can be stated in terms of odds as follows: the odds on a hypothesis in the light of observations are proportional to the probability of those observations given the hypothesis multiplied by the prior odds on this hypothesis. In the present problem, the asyntenic hypothesis remains simple but the syntenic ones must be considered as composite comprising a multitude of sub-hypotheses. There is one of these sub-hypotheses for each pair of values of the map-lengths of adjacent intervals when a given order of the three loci is considered.

From the pedigree data, we calculate on a simple Mendelian model the relative probability of finding these data if there is a true recombination fraction \( \theta_c \) between \( Dm \) and \( Se \); \( \theta_c \) between \( Se \) and \( Lu \); \( \theta_c \) between \( Dm \) and \( Lu \). To have a constant standard for comparison, we standardize these odds (to antilods) by relating them to the probability when \( \theta = \theta_0 = \frac{1}{2} \)-ie, when the loci are on different chromosomes, for example. We shall write the map-lengths corresponding to \( \theta = \frac{1}{2} \) arbitrarily as \( w = 1000M \); they are more correctly infinite—the loci are asyntenic and there is, at least notionally, an infinite amount of crossing-over between them.

We suppose that \( w_i \) (the length of the \( Dm:Se \) interval) yields the recombination fraction \( \theta_0 \); \( w_2 \) (the \( Se:Lu \) map-length) yields \( \theta_0 \); \( w_3 \) (the \( Dm:Lu \) map-length) yields \( \theta_c \). As already mentioned, the map-length and recombination fraction are taken to be related by the Carter-Falconer mapping function of the form: \( 4w = \tan^{-1}2\theta + \tan^{-1}2\theta \). For asynteny, \( \theta_0 = \theta_c = \frac{1}{2} \); the antilod is then 1 by the standardization procedure.

Fortuitously, with one exception in Mohr’s data (1954) to be considered later, no mating segregates fruitfully for more than two of the loci. The antilods from them are therefore largely indifferent to the lengths of the other intervals. Thus, for a given ordering, say \( Dm:Se:Lu \), the antilod (denoted by \( \alpha_{12} \)) for any pair of values of \( w_1 \) and \( w_2 \) (implying \( w_3 = w_1 + w_2 \) is given by

\[
\alpha_{12}(D|w_1,w_2,w_3) = \alpha(D_1|w_1) \cdot \alpha(D_2|w_2) \cdot \alpha(D_3|w_3)
\]

where \( D_1, D_2 \) and \( D_3 \) are non-overlapping parts that constitute the total data, \( D \).

The lods (denoted by \( z \)) are obtained by taking logarithms with base 10:

\[
z_{12}(D|w_1,w_2,w_3) = z(D_1|w_1) + z(D_2|w_2) + z(D_3|w_3).
\]

**Non-independence.** The exceptional mating is in Pedigree 4 of Mohr’s data. The children of I.34 can be scored for recombination between \( Dm \) and \( Se \); \( Se \) and \( Lu \); \( Lu \) and \( Dm \) simultaneously, as she is probably heterozygous. The data, \( D' \), cannot here be split into three non-overlapping parts, each relevant to a single interval. Fortunately, to achieve an approximate adjustment, we can use the following argument:

For a given ordering, say \( Dm:Se:Lu \) (when \( w_1 \) and \( w_2 \) are the primary variables)

\[
\alpha_{12}(D'|w_1,w_2,w_3) = \alpha(D'|w_1) \cdot \alpha(D'|w_2).
\]

To gain a degree of symmetry and hence independence of the ordering we write this as follows:

\[
\alpha_{12}(D'|w_1,w_2,w_3) = [\alpha(D'|w_1) \cdot \alpha(D'|w_2) \cdot \alpha(D'|w_3)] + \alpha(D'|w_3)
\]

or, in logarithmic terms,

\[
z_{12}(D'|w_1,w_2,w_3) = [z(D'|w_1) + z(D'|w_2) + z(D'|w_3)] - z(D'|w_3).
\]
That part of the expression that is enclosed in square brackets is symmetrical in \( w_1, w_2, w_3 \). It is therefore independent of order and is the same in \( z_{13} \) and \( z_{23} \) as in \( z_{12} \). In this particular sibship, the remaining term also happens to have approximately the same numerical value whether it is \( 
abla(z^\prime | w_3) \) for \( z_{12} \) or \( 
abla(z^\prime | w_2) \) for \( z_{13} \) or \( 
abla(z^\prime | w_1) \) for \( z_{23} \). (The interdependence of the coupling phases of the three loci has been ignored but the resulting error is probably trivial in this instance.) It is convenient to rewrite each expression in terms of an adjustment, \( c \), to each lod where \( c \approx -((1/3)\nabla(z^\prime | w_3) - z_3(3,0)/3 \); 
\[\nabla z_{12}(D^\prime | w_1, w_2, w_3) = [\nabla(z^\prime | w_1) + c] + [\nabla(z^\prime | w_2) + c] + [\nabla(z^\prime | w_3) + c].\]

We may paraphrase this argument for events in a single meiosis as follows: For any ordering, the data concerning the two intervals that are contiguous are independent of each other; but the data on the joint interval are entirely determined by the data (from this meiosis) on the other two and must be discounted. In the raw analysis, the offspring of I.34 comprise roughly three recombinants or three non-recombinants—the lods are \( z_l(3,0) \)—for each of the three pairs of loci. As the three orders are taken to be equally likely, the appropriate adjustment involves the discounting of one third of these lods for each of the three pairs. This is approximately achieved by addition of the adjustment, \( c \).

**Posterior Probabilities.** For each ordering of the loci, an appropriate prior density, as discussed later, can be combined with the three antilods for trial values of the true map-lengths, to yield points on a posterior unimodal probability surface. Integration over all possible values of these map-lengths gives the volume under the surface and this represents the posterior unimodal probability of the corresponding synteny hypothesis. Estimates of map-lengths and their limits, conditional on a given sequence of loci, are readily obtained from the appropriate marginal distribution. This procedure of partitioning a hypothesis into a multitude of exhaustive and non-overlapping sub-hypotheses, each with its characteristic pair of map-length values, and then applying Bayes’ argument, is essentially that introduced into the genetic field by Smith (1959).

In order to try to isolate the various components of the calculations, we calculate the ‘change in the odds’ or an ‘averaged antilod’, \( \Lambda_i \), by dividing the posterior integral of a specified part of the distribution by the corresponding prior integral. This averaged antilod is not free of prior elements since we have, in effect, used them as weights in the averaging procedure.

**Detailed Procedure**

**Joint Prior Probability Distribution of Interval Lengths.** Let the total autosomal map length be \( T \) morgans.

Let the length of the \( i \)th autosome be \( A_i > A_{i+1} \) morgans, where \( 0 < i < 22 \).

3-synten. Let the map-lengths FG, GH, FH, between three loci be \( w_1, w_2, w_3 \) morgans respectively. The ordering FGH is called 12 (because the component lengths are \( w_1, w_2 \)); GHF is 31. From equation 4 of Irwin (1955), it follows that that part of the joint probability density for \( w_1, w_2, w_3 \), that concerns the synteny of the loci in a specific ordering on a known autosome, \( i \), is:

\[
f(w_1, w_2, w_3) = 2[A_i - \max(w_1, w_2, w_3)]^{T-3}
\]

if \( 0 \leq \max(w_1, w_2, w_3) \leq A_i \),

\[= 0 \quad \text{if} \quad A_i < \max(w_1, w_2, w_3) \leq A_{i+1}.\]

The combined limits, 0–A1, demarcate the syntenic part of a distribution. (Over all values that lie within these limits and over all three orderings, this part integrates to \( [A_i/T]^3 \)—the prior probability that these autosomal loci, chosen effectively at random, are on this \( i \)th autosome.)

When the loci are syntenic in the specified ordering but the autosome is not specified, the prior density is:

\[
f(w_1, w_2, w_3) = \sum_{i=1}^{22} f(w_1, w_2, w_3)
\]

\[0 \leq \max(w_1, w_2, w_3) \leq A_1.\]

Over all values that lie within the limits 0–A1, the partial distribution described by this density is just a probability function integrates, appropriately, to \( \frac{1}{2} \sum_{i=1}^{22} (A_i/T)^3 \),

the prior probability of autosomal 3-synteny with ordering specified—about 1/900.

For each ordering, \( \max(w_1, w_2, w_3) \) is always the sum of the other two lengths, hence \( f(w_1, w_2, w_3) \) may be presented graphically in a bivariate form, with these two shorter lengths as the primary variables. It constitutes a surface of 22 planes of progressively decreasing tilt as shown in Fig. 1 for the ordering 12. The line of mutual intersection of any two of these is parallel to the line, \( w_1 + w_2 = A_1 \), in which the final plane meets the base. This diagonal line in the base is the limit beyond which the conjunct interval would exceed the length, \( A_1 \), of the longest chromosome (No. 1). As already mentioned, the proportional autosomal map-lengths \( (A_i/T) \) have been taken to be equal to the proportional physical lengths at mitotic metaphase.

We now consider the only other parts of the joint
Fig. 1. The prior probability surface for the ordering $Dm: Se: Lu$ for lengths of two adjacent intervals, $w_1$ and $w_2$, on an unknown autosome. The surface consists of 22 planes of decreasing tilt intersecting each other in lines parallel to the line $w_1 + w_2 = A_1$, where $A_1$ is the map length of No. 1 chromosome (the longest). The surface represents only the 3-syntenic part of the distribution of map-interval lengths.

distribution of map-lengths that have non-zero prior densities. These parts involve asynten but we may avoid the introduction of a two-valued 'tenic' variable by adopting the convention that an interval between asyntenic loci is ascribed a length of 1000 morgans except when the additivity rule of map-lengths requires it to be longer. We may partially justify this device. In a formal sense, the independent assortment of loci on different pairs of autosomes produces the same proportion of gametes that are recombinant for these loci as would be produced by an infinite amount of crossing-over between syntenic loci. The interval between chromosome pairs might indeed be regarded as one of infinite susceptibility to 'crossing-over' so that physical continuity was constantly being disrupted. Formally, then, the map-length between asyntenic loci is infinite but there are important advantages in substituting an arbitrarily large value such as 1000 morgans: addition of map-lengths is, for one thing, still meaningful.

2-syteny and its part of the distribution. There are again three possible orderings of the loci, here depending on which one is asyntenic. This one will be written as the last locus: for example, the ordering 12 (FGH or GFH) then implies that F and G are syntenic, H is not. $\text{Min}(w_1, w_2, w_3)$ denotes the map distance between the two syntenic loci. For a specific ordering, the joint prior distribution of map-lengths has a contribution, $f_i(w_1, w_2, w_3)$, that arises when the syntenic loci are on the $i$th autosome:

$$f_i(w_1, w_2, w_3) = \begin{cases} 2[A_i - \text{min}(w_1, w_2, w_3)]T^{-2}(T - A_i)T^{-1} & \text{if } 1000 \leq \text{max}(w_1, w_2, w_3) \leq 1000 + A_i \\ 0 & \text{if } 1000 + A_i < \text{max}(w_1, w_2, w_3) \end{cases}$$

(For this specific ordering and for this $i$th chromosome and over all values that lie within either of these sets of limits, the corresponding part of the distribution that is described by this density function integrates appropriately to $[A_i/T]^2 - [A_i/T]^3$. This is the probability that two specific loci lie on the $i$th autosome and that the third lies elsewhere.)
When the autosome on which the two loci are syntenic is not specified (the usual case), the prior density is:

\[ f(w_1, w_2, w_3) = \sum_{i=1}^{22} f_i(w_1, w_2, w_3). \]

That part of the distribution for which this density is appropriate integrates to \( \sum_{i=1}^{22} [(A_i/T)^2 - (A_i/T)^3] \).

Its numerical value is about 46/900. The full distribution includes of course three such parts, one for each 'ordering' specifying which locus is syntenic.

Asynteny of all three loci. The probability for asynteny is about 759/900 as already mentioned and is not distributed.

**Averaged Antilod and Posterior Distribution.** For a specific ordering of the loci, one interval length is completely defined by the sum of the other two, hence it does not make any non-redundant contribution to the prior densities. The data D may, however, refer to all three intervals largely in an independent and non-overlapping manner and, accordingly, may be considered in three parts—\( D_1, D_2, \) and \( D_3 \). The posterior probability density (unnormed) for interval lengths, \( w_1, w_2, w_3 \), in ordering 12 (where \( w_3 = w_1 + w_2 \) is therefore the algebraic product of the three likelihood ratios, \( \alpha(D_1|w_1), \alpha(D_2|w_2), \alpha(D_3|w_3) \), and the prior joint density \( f_{12}(w_1, w_2, w_3) \). The above presentation differs superficially from that of Bolling (1970) which is in terms of *conditional* distributions.

**Se Linkage Group: An Illustration of the Method.** The lods were computed for the \( Dm: Se, \) \( Dm: Lu, \) and \( Se: Lu \) intervals from all the available published and unpublished pedigree data by the program of Renwick and Schulze (1961). They are given in Table I of Renwick et al. (1971)—for the equal-susceptibility situation, i.e., when male and female map-lengths are taken to be equal.

For a specific ordering of the loci, the volume of the 3-syntenic part of the posterior probability distribution (unnormed) was determined from these three sets of total lods by the MAPIN computer program (Bolling, 1970). MAPIN is written in Fortran IV and uses two numerical integration methods which usually show good agreement with each other. The ratio of this volume to the prior volume is the averaged antilod, \( A' \), or the change in the odds, for this particular ordering of the loci. The procedure was repeated for the other possible orderings.

In Table I, these values of \( A' \) obtained from Mohr's data (1954) are set out for each synteny hypothesis. Those for the 2-synteny hypotheses

**TABLE I**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Loci</th>
<th>Prior Odds</th>
<th>Change in Odds</th>
<th>Posterior Odds</th>
<th>Posterior Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-syntenic</td>
<td>( Dm: Se; Lu )</td>
<td>1</td>
<td>( 3 \times 10^{10} )</td>
<td>( 3 \times 10^{10} )</td>
<td>6.8 \times 10^{10}</td>
</tr>
<tr>
<td>3-syntenic</td>
<td>( Dm: Lu; Se )</td>
<td>1</td>
<td>( 2 \times 10^{10} )</td>
<td>( 2 \times 10^{10} )</td>
<td>6.8 \times 10^{10}</td>
</tr>
<tr>
<td>3-syntenic</td>
<td>( Lu; Dm; Se )</td>
<td>1</td>
<td>( 1 \times 10^{10} )</td>
<td>( 1 \times 10^{10} )</td>
<td>6.8 \times 10^{10}</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Dm: Se; Lu )</td>
<td>46</td>
<td>6</td>
<td>( 14 )</td>
<td>( 14 )</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Dm: Lu; Se )</td>
<td>46</td>
<td>14</td>
<td>( 267 ) only</td>
<td>( 264 ) only</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Lu; Dm; Se )</td>
<td>759</td>
<td>1</td>
<td>( 759 ) only</td>
<td>( 759 ) only</td>
</tr>
<tr>
<td>Any of these . . .</td>
<td>900</td>
<td>. . .</td>
<td>( 7 \times 10^{10} )</td>
<td>( 7 \times 10^{10} )</td>
<td>1.00</td>
</tr>
</tbody>
</table>

**TABLE II**

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Loci</th>
<th>Prior Odds</th>
<th>Change in Odds</th>
<th>Posterior Odds</th>
<th>Posterior Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-syntenic</td>
<td>( Dm: Se; Lu )</td>
<td>1</td>
<td>( 30 \times 10^{10} )</td>
<td>( 30 \times 10^{10} )</td>
<td>55 \times 10^{10}</td>
</tr>
<tr>
<td>3-syntenic</td>
<td>( Lu; Dm; Se )</td>
<td>1</td>
<td>( 20 \times 10^{10} )</td>
<td>( 20 \times 10^{10} )</td>
<td>55 \times 10^{10}</td>
</tr>
<tr>
<td>3-syntenic</td>
<td>( Lu; Se; Dm )</td>
<td>1</td>
<td>( 4 \times 10^{10} )</td>
<td>( 4 \times 10^{10} )</td>
<td>6.6 \times 10^{10}</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Dm: Se; Lu )</td>
<td>46</td>
<td>23</td>
<td>( 1058 ) only</td>
<td>( 1058 ) only</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Dm: Lu; Se )</td>
<td>46</td>
<td>14</td>
<td>( 644 ) only</td>
<td>( 644 ) only</td>
</tr>
<tr>
<td>2-syntenic</td>
<td>( Lu; Dm; Se )</td>
<td>759</td>
<td>1</td>
<td>( 759 ) only</td>
<td>( 759 ) only</td>
</tr>
<tr>
<td>Any of these . . .</td>
<td>900</td>
<td>. . .</td>
<td>( 56 \times 10^{10} )</td>
<td>( 56 \times 10^{10} )</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The odds on the 3-syntenic orderings, \( Dm: Se; Lu, Lu: Dm; Se, Dm: Lu; Se \), are seen to be left at 30:20:4 or about 6:4:1.
were computed, for strict comparability, on the 3-
synteny model with the invention of a dummy as an
extra syntenyic locus on which the data do not bear.
(All orderings were, of course, again considered.)
Mohr's data alone yielded final odds on 3-synteny
of 12:1 over all other hypotheses. With the inclusion
of the data of Renwick et al (1971), these odds
became 92:1 (Table II), corresponding to a post-
terior probability of 0.99 that the dystrophia

\[ P(R) = \int_0^{0.5} \theta p(\theta).d\theta = 0.073. \]

locus belongs to the Se:Lu linkage group. Given
3-synteny, the current odds on the alternative
sequences, Dm: Se: Lu, Dm: Lu: Se, Se: Dm: Lu,
are found to be 6:4:1.

The marginal distributions were also computed
by MAPIN. That for the longest interval was
constructed by integrating the distribution along
lines parallel to \( w_1 + w_2 = A \) for the ordering 12;
on any equivalent lines otherwise. As expected,
the shape of the marginal distribution for one of
the intervals—the Lu:Se interval—is so dominated
by the data bearing directly on it that it remains
almost invariant whatever the ordering of the loci,
whereas the marginal distributions for the Dm: Se
and Dm: Lu interval lengths reflect fewer data and
therefore vary considerably among the three possi-
ble orderings.

The aggregate of the three marginal distributions
for the Dm: Se interval gives a most probable map-
length of 0.04 morgans between the two loci. The
distribution of the corresponding recombination
fraction, \( \theta \) (see Fig. 2) is critical when we apply
these data in prenatal diagnosis as explained later.

A joint analysis (such as that outlined in this
paper) is required not only to estimate each interval-
length using all the data but to assess too the
probability that Dm does indeed belong to the
Se: Lu linkage group. This follows from the mu-
tuality of the lengths of the three intervals, one
being the sum of the other two. The powerful
influences of the prior probabilities are, in them-
selves, good reasons for this joint analysis being
of Bayesian type.

**Antenatal Prediction of Myotonic Dystrophy.**

It was mentioned in the introduction that link-
geages can theoretically be used for indirect diag-
nosis in *utero*. The dystrophia:secretor linkage is one of
the first autosomal examples. An important re-
quirement has been met by the recent finding by
Harper and Hutchinson (1970) that the amniotic
fluid does reflect the secretor genotype of the fetus.
(That study was stimulated by our analysis of Mohr's
data.) In certain suitable families in which the
affected parent is heterozygous for both loci, it
should be possible to infer what allele was trans-
mited from the dystrophy locus from knowledge of
what allele was transmitted from the secretor locus.
The security of the inference would, of course,
depend on the risk that the gamete was recombinant.
This risk \( P(R) \) would be equal to the true re-
combination fraction \( \theta \). This remains unknown but we
do possess in the light of the linkage data a prob-
ability distribution for the values it might take.
We cannot justifiably take the modal estimate of \( \theta \)
place of its true value because the distribution is
skewed markedly (towards high values of \( \theta \)).
Pursuing such reasoning, we take the posterior
probability, \( P(R) \), that a particular gamete will be
recombinant for alleles at the dystrophy and
secretor loci to be the sum of elements of condi-
tional probability, \( p(R|\theta) \cdot d\theta \), each weighted by
the posterior probability density, \( p(\theta) \), appro-
appropriate for recombination fractions lying
between \( \theta \) and \( \theta + d\theta \).

For sufficiently small increment, \( d\theta \),

\[ P(R) = \int_0^{0.5} p(R|\theta) \cdot p(\theta).d\theta = \int_0^{0.5} \theta p(\theta).d\theta. \]

We require, therefore, the centre of gravity (ie,
the mean) of the posterior distribution of \( \theta \). We
can obtain it graphically by plotting the \( \theta p(\theta) \cdot d\theta 
function against \( \theta \) and integrating it. To do this,
we measure the area under it and the area under
the distribution itself. After allowance for scaling
(such as the 20-fold scaling used in Fig. 2) the ratio

---

**Fig. 2.** Posterior distribution, \( p(\theta) \cdot d\theta \), of the recombination fraction, \( \theta \), between Dm and Se, summed over all three orders of the loci, Dm, Se, Lu. The scale is unNormed. This distribution is derived from the marginal distribution of the corresponding map interval.

---

Renwick and Bolling
of the areas which is found to be about 0.073 is the required probability if, on the one hand, there is synteny of Dm and Se (probability 0.99). If, on the other hand, this condition of synteny does not hold (probability 0.01), the recombination fraction is exactly 0.5. Using the combined data of the present work and of Mohr (1954), we therefore assess the probability that any specified gamete is a recombinant as 0.99(0.073) + 0.01(0.5) or 0.078. The probability that it is a non-recombinant is thus about 92%—so a prediction, that could be given with 100% security but for recombination, has about 92% security when we take recombination into account.

It might seem, at first sight, that the probability of a recombinant, contrary to the above, should be equal to the probability of recombination, as estimated modally from the data. If the estimate of \( \theta \) were associated with only a trivial bias or only a trivial variance this would indeed be approximately true. In more real situations, all possible values of \( \theta \) must be considered with weights proportional to their posterior probability densities. This can perhaps be better understood by means of an artificial oversimplification:

Consider that only three values, 0.05, 0.1, 0.2 are possible with posterior probabilities 0.7, 0.2, 0.1 respectively. The most probable recombination fraction would then be 0.05 but the probability, \( P(R) \), of a particular gamete being a recombinant would be

\[
P(R) = (0.7)(0.05) + (0.2)(0.1) + (0.1)(0.2) = 0.075.
\]

**Influence of Sex.** The Se : Lu region is known to be one of those that have an excess of crossing-over in females. If this region and its neighbours are average in this respect, we could perhaps use, for its female: male ratio of susceptibility to crossing-over, the overall estimate, 1:40, that we obtained recently from an unpublished analysis of available data on a number of autosomal linkages. To a first approximation, we can then derive map-length and recombination estimates for males by multiplying the above raw estimates by 5/6 and for females by multiplying by 7/6. The security of prediction is then, for a father, \( 100 - (7.8)(5/6) \) or 93% and, for a mother, \( 100 - 7.8(7/6) \) or 91%.

The main approximation involved here lies in the fact that the raw estimates are based partly on data relevant to male meiosis and partly on data relevant to female meiosis. These two sets of data can be expected in general to be unequal in size and moment, so that raw estimates obtained by pooling them are only approximations to the required unweighted average of male and female estimates separately derived. A more sophisticated analysis may be envisaged. It would be possible to compute these estimates subject to the current estimate of the overall susceptibility ratio if this ratio were deemed approximately uniform and therefore appropriate to this chromosomal region.

**Summary**

Re-analysis of data of Thomasen (1948) and Mohr (1954) that was undertaken as part of a wider computer-based linkage analysis gave evidence that the loci of myotonic dystrophy (Dm), ABH secretion (Se), and the Lutheran blood group system (Lu) might be on the same chromosome (syntenic). Like Mohr, who made the same observation by a different method, we should have found the evidence difficult to evaluate but for a fortuitous concurrence with the designing of a Bayesian computer analysis for linkage data involving 3 or 4 loci. Extension to 5 loci appears feasible but extension to 6 without further approximation does not. That method is published for the first time here.

The posterior probability, 0.93, that the 3 loci are syntenic stimulated collection of further data that now raise it to 0.99 (Renwick et al, 1971). The method when applied to these two sets of data, gives a preferred map Dm : Se : Lu with consecutive intervals of lengths 0.04:0.13 morgans, though other orderings are also possible.

The Dm : Se linkage is important since it can be used for an indirect and novel form of prenatal diagnosis. In suitable pedigrees, the dystrophia can be predicted with 92% security on the basis of the secretor phenotype of the fetus as determined on its amniotic fluid (Harper and Hutchinson, 1970).

Drs E. A. Murphy and Helen Abbey, and Professors M. J. M. Bernal and P. Armitage made material contributions to the development of the method. The work was carried out chiefly in London in 1969 during part of the tenure by D.R.B. of a training award from the US Public Health Service. It was supported also by grants G968206B from the Medical Research Council and GM-10189 from the National Institutes of Health (Dr V. A. McKusick, Johns Hopkins Hospital Medical School, Baltimore). The program was developed on the CDC 6600 computer of the University of London Computing Centre through the Birkbeck terminal.

**References**


Renwick and Bolling


