Chapter 5. Hypotheses Tests for Pooling Band Recovery Data Sets

In Chapters 2 and 3 we dealt at length with the analysis of individual data sets. In the context of adult banding, it is straightforward that an individual data set is the band recovery data from k years of banding for ℓ years of recovery (see Section 2.1). If both adults and young are banded (Section 3.1), the two band recovery arrays are considered as one data set. In either instance, the banding is typically done in the same relatively small area each year, and the sex of the birds is often determined. Consequently, in any large-scale analysis of band recoveries we are likely to have numerous data sets representing both sexes and a variety of banding areas. In this chapter we discuss several tests developed to aid in making decisions about pooling such data sets. Basically, such tests examine the question of whether or not the data sets have the same underlying survival and recovery rate parameters.

Initially, the motivation for this type test came from the issue of combining the two data sets for adult males and females banded in the same area. This test for sex differences is mathematically the same as the test for whether two data sets from different areas (for the same sex) have the same underlying parameter values. For example, if adult male mallards had been banded each winter in northeastern Colorado and western Nebraska, we may wish to pool these data for analysis. Whereas there are only two sexes, we may have data from more than two areas. The logical extension of these tests has been developed for examining questions of pooling data sets over several areas.

In all of these tests, the null hypothesis is that the data sets are described by the same survival and recovery rate parameters and hence could be pooled. Rejection of this hypothesis is evidence that the data sets should be analyzed separately and then the estimates may be averaged, rather than pooling the data. If a test is not significant, say at the 5% level, we have evidence that the data sets may be pooled for analysis.

5.1 Testing Equality of Survival and Recovery Rates for Adult Data Sets

Testing Equality of Adult Male and Female Parameters

As far as possible, the notation of Chapter 2 will be used in this section; specifically, the reader is referred to Sections 2.1 and 2.2.

For this test to be meaningful, the data sets for males and females must come from the same banding study, more specifically, from the same area and cover the same years of banding and recovery. Under Model 1 let the parameters for the two data sets be

S_{im} = adult male survival rate in year i,
S_{if} = adult female survival rate in year i,
f_{im} = adult male recovery rate in year i,
f_{if} = adult female recovery rate in year i.

The null hypothesis is $S_{im}=S_{if}$, $i=1,\ldots,\ell-1$ and $f_{im}=f_{if}$, $i=1,\ldots,\ell$. Here the general case of k years of banding and ℓ years of recoveries is assumed (hence $\ell > k$ is possible).

The alternative hypothesis is a composite one where: (1) only recovery rates differ, (2) only survival rates differ, or (3) both recovery and survival rates differ. If the null hypothesis is rejected, we do not know where the differences in parameter values lie. The way to examine this is to analyze each data set separately by program ESTIMATE and examine the resultant parameter estimates.

From Section 2.1 the following summary statistics were defined for an individual data set:

R_i = total of row i of the data array, $i=1,\ldots,k$,
C_i = total of column i of the data array, $i=1,\ldots,\ell$,
T_i = a block total $i=1,\ldots,\ell$.

Define R_{im}, C_{im}, and T_{im} as these summary statistics for the set of male recovery data, and similarly define R_{if}, C_{if}, and T_{if} for the female recovery data. The test of the null hypothesis of no difference in male and female parameters is based on these statistics, and takes the form of $k+\ell-1$ separate chi-square test statistics, each with 1 df, computed on separate 2 by 2 contingency tables.

Defining the sample sizes for males and females as N_{im} and N_{if}, the set of contingency tables is

R_{im}	$N_{im}-R_{im}$
R_{if}	$N_{if}-R_{if}$

$, i=1,\ldots,k$

and

C_{im}	$T_{im}-C_{im}$
C_{if}	$T_{if}-C_{if}$

$, i=1,\ldots,\ell-1.$

There are $k+\ell-1$ of these tables. For each one, the usual chi-square goodness of fit statistic is computed. The first set of k, 2 by 2 tables uses the row sums and sample sizes from the data, while the second set ($\ell-1$ tables) is based on the column and block totals. For triangular data arrays, $\ell=k$.

For each of these $k+\ell-1$ tables one chi-square test statistic is computed. These values are then added to obtain one test statistic which has a chi-square distribution with $k+\ell-1$ degrees of freedom under the null hypothesis. If this test value exceeds the critical level, at the 5% significance level for instance, for a chi-square variable with $k+\ell-1$ degrees of freedom, then one rejects the null hypothesis. In this event pooling the data sets is not justified.

An Example of the Male-Female Test

This test is available as an option within the FORTRAN program BROWNIE (see Chapter 6 for details of how to input data and obtain this test).

The test of the null hypothesis is easy to use and interpret as we illustrate in Example 5.1. Data on male and female mallards banded during the winter in Illinois, 1963-70, are used to illustrate the testing procedure (the data for adult males is the same as that used in Example 2.6). The computer output, shown in Example 5.1, displays the input data, and then various summary statistics. The notation for these various subtotals (summary statistics) used by the program differs from this chapter in that RROW(I), RCOL(I), T(I) correspond to R_{im}, C_{im}, T_{im}, respectively, and QROW(I), QCOL(I), U(I) correspond to R_{if}, C_{if}, T_{if}, respectively. Finally the contingency tables and individual chi-square values are given, followed by the total chi-square which is the test statistic. The program also prints the achieved significance level of the test statistic, i.e., the probability of getting a value as large as that observed, if the null hypothesis is true.

In Example 5.1 we see that the total chi-square value is 314.17 with 18 df. The probability of a value this large, if the null hypothesis of no difference in male-female parameter values is true, is essentially zero (to 8 decimal places). Therefore we reject the null hypothesis and conclude that recovery and/or survival rates differ significantly by sex, and the data sets must be analyzed separately. From the separate analyses we can compare male and female recovery and survival rates to determine where the differences are.

The program shows the individual contingency tables for the triangular part of the data arrays under the titles MATRIX 1 and MATRIX 2. MATRIX 3 relates to the nontriangular portion of the data ($s>0$), and is not shown in detail. The output labeled MATRIX 1 gives those tables based on row totals. For example, in 1963 ($i=1$) there were 2,583 males banded. A total of 279 recoveries were recorded from this banded cohort. Similarly, for females $N_{1f}=1{,}478$ and $R_{1f}=94$. Thus the first 2 by 2 table is

279	2,583 − 279
94	1,478 − 94

$\equiv$

279	2,304
94	1,384

The individual chi-square statistic from this table is 22.232.

Example 5.1

MALE AND FEMALE MALLARDS BANDED DURING THE WINTER IN ILLINOIS, 1963-70

ADULT MALE INPUT MATRIX

1963	2583.	91.	89.	24.	18.	16.	11.	8.	7.	7.	2.	6.
1964	3075.	0.	141.	45.	52.	50.	17.	30.	21.	16.	7.	3.
1965	1155.	0.	0.	27.	31.	21.	8.	19.	7.	9.	4.	3.
1966	3418.	0.	0.	0.	156.	92.	44.	50.	49.	34.	23.	5.
1967	3100.	0.	0.	0.	0.	113.	68.	57.	65.	41.	23.	10.
1968	2400.	0.	0.	0.	0.	0.	63.	52.	59.	44.	30.	12.
1969	2601.	0.	0.	0.	0.	0.	0.	91.	80.	58.	37.	25.
1970	4433.	0.	0.	0.	0.	0.	0.	0.	222.	169.	95.	46.

ADULT FEMALE INPUT MATRIX

1963	1478.	40.	31.	8.	11.	2.	0.	2.	0.	0.	0.	0.
1964	1525.	0.	72.	20.	15.	7.	5.	1.	2.	1.	0.	0.
1965	319.	0.	0.	8.	7.	3.	0.	1.	3.	1.	0.	0.
1966	1805.	0.	0.	0.	63.	27.	14.	5.	5.	2.	2.	2.
1967	1400.	0.	0.	0.	0.	39.	14.	17.	10.	7.	5.	2.
1968	900.	0.	0.	0.	0.	0.	17.	9.	15.	10.	6.	1.
1969	1400.	0.	0.	0.	0.	0.	0.	39.	21.	10.	10.	0.
1970	1789.	0.	0.	0.	0.	0.	0.	0.	63.	39.	11.	4.

BASIC SUBTOTALS

I	RROW(I)	RCOL(I)	CROW(I)	QCOL(I)	T(I)	U(I)	W(I)	Z(I)
1	279.00	91.00	94.00	40.00	279.00	94.00	91.00	0.0
2	382.00	230.00	123.00	103.00	570.00	177.00	261.00	188.00
3	129.00	96.00	23.00	36.00	469.00	97.00	124.00	363.00
4	453.00	257.00	120.00	96.00	826.00	181.00	290.00	419.00
5	377.00	292.00	94.00	78.00	946.00	179.00	331.00	597.00
6	260.00	211.00	58.00	50.00	914.00	159.00	244.00	700.00
7	291.00	307.00	80.00	74.00	994.00	189.00	342.00	771.00
8	532.00	510.00	117.00	119.00	1219.00	232.00	566.00	761.00
					709.00	113.00		768.00

CONTINGENCY CHI-SQUARE TEST FOR DIFFERENCES DUE TO SEX

MATRIX 1

	2 X 2 CONTINGENCY TABLE		CORRESPONDING CHI-SQUARE STATISTIC WITH 1 DEGREE OF FREEDOM
I= 1	279	2304	
	94	1384	22.232
I= 2	382	2693	
	123	1402	19.803
I= 3	129	1066	
	23	296	3.583
I= 4	453	2965	
	120	1685	52.763
I= 5	377	2723	
	94	1306	30.535
I= 6	260	2140	
	58	842	14.479
I= 7	291	2310	
	80	1320	32.413
I= 8	532	3901	
	117	1672	40.685

MATRIX 2

	2 X 2 CONTINGENCY TABLE		CORRESPONDING CHI-SQUARE STATISTIC WITH 1 DEGREE OF FREEDOM
I= 1	91	188	
	40	54	3.047
I= 2	230	340	
	103	74	17.401
I= 3	96	373	
	36	61	12.452
I= 4	257	569	
	96	85	31.348
I= 5	292	654	
	78	101	11.014
I= 6	211	703	
	50	109	5.144
I= 7	307	687	
	74	115	4.972
I= 8	510	709	
	119	113	7.096

MATRIX 3

CONTINGENCY TABLE FOR S GREATER THAN 1

I= 9

110.221.378.
9. 34. 70.

5.21 WITH 2 D.F.

TOTAL CHI-SQUARE	314.17 WITH 18 D.F.

PROBABILITY OF A CHI-SQUARE VALUE LARGER THAN 314.17 = 0.0

Under the heading MATRIX 2, the 2 by 2 tables based on C_i and T_i are given for the triangular part of the data array. For example, $C_{1m}=91$, $T_{1m}=279$, $C_{1f}=40$, and $T_{1f}=94$. Hence for $i=1$, the table is

91	279 − 91
40	94 − 40

≡

91	188
40	54

When using this test, one should look at the individual chi-square values on 1 df, as well as the overall result (314.17, with 18 df in this example). If the overall test strongly rejects the hypothesis of no difference ($P<0.005$), one can safely conclude that the data sets should not be pooled. But in the event the result is significant at lesser levels like ($0.05<P<0.005$), one should examine the separate chi-square tests on the different contingency tables. It may well be that just one of these individual chi-square values is quite large, hence "causing" the significant overall test result. In this event it is probably the male and female recovery rates in 1 year that are different, and other parameters may be very similar by sex. Pooling the data is reasonable in this case. If, however, no one single chi-square stands out, and they all tend to be a little larger than expected under the null hypothesis, then it is probably because a slight but consistent difference exists in male and female parameters and the data should not be pooled.

This test for differences in adult male and female parameters was developed by Brownie (1973); it is the only one of its kind that is programmed for computer computation. A similar test is described below for adult and young data; also both these tests are extended to the question of pooling data for the same sex over different areas.

Testing Adult Data Sets for Pooling Over Areas

Nothing in the theory of the test described above is specific for testing for sex specificity. Although this was the motivation of the test, it is nothing more than a test of the null hypothesis that two adult band recovery data sets (having the same values of k and ℓ) have the same survival and recovery rates. It is logical to apply this type test to the question of pooling data sets across areas for the same sex, species, and banding and recovery years. For two areas, the above test suffices. For more than two areas, the test has a simple extension.

For r areas the appropriate test for pooling is based on a series of r by 2 contingency tables. As before, there will still be exactly $k+\ell-1$ tables. The null hypothesis is that Model 1 with the same annual survival and recovery rates fits all r data sets (hence they can validly be pooled).

Instead of indexing data sets by m and f, let the data and parameters have a second index for area, $a=1,\ldots,r$. For example, the row total R_i from data set a becomes R_{ia}. The set of statistics R_i, C_i, and T_i, as well as N_i, must first be tabulated for each data set. The following r by 2 contingency tables are then easily written:

R_{i1}	$N_{i1}-R_{i1}$
R_{i2}	$N_{i2}-R_{i2}$
.	.
.	.
.	.
R_{ir}	$N_{ir}-R_{ir}$

$, i=1,\ldots,k.$

C_{i1}	$T_{i1}-C_{i1}$
C_{i2}	$T_{i2}-C_{i2}$
.	.
.	.
.	.
C_{ir}	$T_{ir}-C_{ir}$

$,i=1,\ldots,\ell-1.$

Note that each row comes from a different data set (hence different area), but in a given type of table, each row is otherwise the same. Thus the extension aspect of this test beyond the case of $r=2$ is simply to add more rows onto each basic table, each row being computed in the same way (for a given table) from a different data set.

Once these $k+\ell-1$ tables have been developed, the test is straightforward. The usual chi-square test statistic for homogeneity is computed for each table. These test statistics are each approximately chi-square distributed with $r-1$ degrees of freedom under the null hypothesis. One then adds all $k+\ell-1$ test statistics. The result is the overall test statistic for testing the null hypothesis that all data sets have the same recovery and survival rate parameters. The test statistic is distributed as chi-square with $(k+\ell-1)\,(r-1)$ degrees of freedom when the null hypothesis is true. If this test value exceeds the critical value for the chosen significance level, we reject the null hypothesis and conclude that pooling of data sets is not warranted.

5.2 Testing Adult and Young Data Sets for Pooling

We will give the details of this test for comparing r data sets; these could be on the same species and sex for different areas, or if $r=2$, they could be data sets for one area for males and females. This test is not programmed in BROWNIE and it would have to be done by hand or a program written. To compute the test, a set of summary statistics must first be computed for each data set. These are in the notation of Chapter 3,

$$R_{i\cdot}\ \text{(adult row totals)} \qquad ,i=1,\ldots,k,$$

$$Q_{i\cdot}\ \text{(young row totals)} \qquad ,i=1,\ldots,k,$$

$$Q_{ii}\ \text{(first year recoveries for young)} \qquad \begin{matrix} ,i=1,\ldots,k-1 \text{ if } \ell=k \\ ,i=1,\ldots,k \quad \text{if } \ell>k \end{matrix}$$

$$W_i=\begin{cases} R_{\cdot i}+Q_{\cdot i}-Q_{ii} & ,i=1,\ldots,k \text{ if } s\leq 1 \\ R_{\cdot i}+Q_{\cdot i} & ,i=k+1,\ldots,k+s-1 \text{ if } s>1 \end{cases}$$

$$X_i=\begin{cases} T_i+U_i-Q_{i\cdot} & ,i=1,\ldots,k \text{ if } s\leq 1 \\ T_i+U_i & ,i=k+1,\ldots,k+s-1 \text{ if } s>1. \end{cases}$$

The statistics $R_{\cdot i}$, $Q_{\cdot i}$ (column totals), T_i and U_i (block totals) would have to be computed as intermediate values. Program BROWNIE computes and prints these statistics, except for X_i, for the triangular part of the data array. Thus, by analyzing each data set first, most of the computation is done by the computer.

Given these summary statistics, for each data set, one then constructs a series of r by 2 contingency tables. In each table the entries are of the same type but each row comes from a different data set.

From the statistics $R_{i\cdot a}$, construct the r by 2 contingency tables:

data set

1	$R_{i\cdot 1}$	$N_{i1}-R_{i\cdot 1}$
2	$R_{i\cdot 2}$	$N_{i2}-R_{i\cdot 2}$
.	.	.
.	.	.
.	.	.
r	$R_{i\cdot r}$	$N_{ir}-R_{i\cdot r}$

$, i=1,\ldots,k.$

In the above, the added subscript to the notation denotes the data set from which the values came. Thus, N_{ia} is the number of adults banded in year i, and area (or sex) a. Similarly, $R_{i\cdot a}$ is the row total of recoveries from adults banded in year i in area a (or of sex a). This augmented notation will also be used below.

The additional contingency tables to construct are:

$Q_{i\cdot 1}$	$M_{i1}-Q_{i\cdot 1}$
$Q_{i\cdot 2}$	$M_{i2}-Q_{i\cdot 2}$
.	.
.	.
.	.
$Q_{i\cdot r}$	$M_{ir}-Q_{i\cdot r}$

$, i=1,\ldots,k$

Q_{ii1}	$Q_{i\cdot 1}-Q_{ii1}$
Q_{ii2}	$Q_{i\cdot 2}-Q_{ii2}$
.	.
.	.
.	.
Q_{iir}	$Q_{i\cdot r}-Q_{iir}$

$$, \begin{cases} i=1,\ldots,k-1 & \text{if } \ell=k \\ i=1,\ldots,k & \text{if } \ell>k \end{cases}$$

W_{i1}	$X_{i1}-W_{i1}$
W_{i2}	$X_{i2}-W_{i2}$
.	.
.	.
.	.
W_{ir}	$X_{ir}-W_{ir}$

$, i=1,\ldots,\ell-1.$

If $s=0$ $(\ell=k)$, there are $4k-2$ tables (each r by 2); if $s>0$, there are $4k+s-1$ tables.

For each contingency table one computes the usual chi-square test of homogeneity. Each of these individual test statistics is (approximately) distributed as chi-square with $r-1$ degrees of freedom under the null hypothesis that the H_1 model with the same survival and recovery rates fits the data from all areas. The test statistic for this null hypothesis is the sum of all these individual chi-square values. Under the null hypothesis this sum is distributed as a chi-square variable with degrees of freedom

$$\begin{aligned}&(4k-2)\,(r-1)\,, \text{ if } s=0\\ &(4k+s-1)\,(\mathrm{r}-1), \text{ if } s>0\,.\end{aligned}$$

One rejects the null hypothesis if the test statistic value exceeds the critical value for the chosen significance level.

5.3 Mathematical Background

The reader with little knowledge of mathematics can skip this section. Of the above tests, only the test for differences in adult male and female parameters has previously appeared in the literature (Brownie 1973). Consequently, this section will supply the basic mathematical background of all these tests.

Adult Data

Under Model 1 for adult data a minimal sufficient statistic (MSS) is

$$\mathcal{S}=\{R_1,\ldots,R_k,C_1,\ldots,C_{\ell-1}\}\,.$$

Because of the recursive relationship

$$\begin{aligned}&T_1=R_1\,,\\ &T_{i+1}=T_i-C_i+R_{i+1} \qquad ,\ i=1,\ldots,k-1\end{aligned}$$

and if $s>0$,

$$T_{k+j}=T_{k+j-1}-C_{k+j-1} \qquad ,j=0,\ldots,s-1\,,$$

the T_i are implicitly part of the minimal sufficient statistic $\mathcal{S}$ (T_i becomes a shorthand notation for a function of the MSS).

As determined initially by Robson and Youngs (1971) the probability distribution of $\mathcal{S}$ is given by

$$R_i \sim B(N_i,\rho_i) \qquad ,i=1,\ldots,k$$

and

$$C_i \text{ given } T_i \sim B(T_i, f_i/\rho_i) \qquad ,i=1,\ldots,\ell-1\,.$$

These distributions are mutually independent. In this notation $B(n, p)$ is the binomial distribution, and from Chapter 2

$$\rho_i=f_i+S_if_{i+1}+\cdots+S_i\cdots S_{\ell-1}f_\ell\,.$$

It is thus straightforward to write down the probability distribution of $\mathcal{S}$, $P\{\mathcal{S}\}$, under Model 1.

Let $\mathcal{S}_m$ and $\mathcal{S}_f$ be the minimal sufficient statistic for separate adult male and female recovery data under the alternative hypothesis of different parameter rates. Under the null hypothesis, H_0, of equal parameters, the MSS is

$$\mathcal{S}_0=\{R_{im}+R_{if},\, i=1,\ldots,k,\, C_{im}+C_{if},\, i=1,\ldots,\ell-1\}\,.$$

A test of H_0 can be based on the probability distribution of the MSS under the alternative hypothesis, conditional on $\mathcal{S}_0$, and given H_0 is true. Symbolically we need to find

$$\mathrm{P}_{H_0}\{\mathcal{S}_m,\ \mathcal{S}_f \mid \mathcal{S}_0\}\,.$$

Using the above results on the distribution of $\mathcal{S}$ and the independence of the two data sets leads to the result

$$\mathrm{P}_{H_0}\{\mathcal{S}_m,\ \mathcal{S}_f \mid \mathcal{S}_0\}=\prod_{i=1}^{k}\frac{\binom{N_{im}}{R_{im}}\binom{N_{if}}{R_{if}}}{\binom{N_{im}+N_{if}}{R_{im}+R_{if}}}\prod_{i=1}^{\ell-1}\frac{\binom{T_{im}}{C_{im}}\binom{T_{if}}{C_{if}}}{\binom{T_{im}+T_{if}}{C_{im}+C_{if}}}\,.$$

This is a product of $k+\ell-1$ independent hypergeometric distributions, from which the series of 2 by 2 contingency tables follow as a basis for a chi-square test of H_0. This same approach is used to derive the other tests of this chapter.

The parameters appearing in the distribution of the MSS (i.e., ρ_i and f_i/ρ_i) are a one-to-one transformation of the basic parameters S_i and f_i. We are testing the null hypotheses that $\rho_{im} \equiv \rho_{if}$, $i=1,\ldots,k$ and $(f_{im}/\rho_{im}) \equiv (f_{if}/\rho_{if})$, $i=1,\ldots,\ell-1$. Each individual equality has its own chi-square test. For each contingency table we can determine exactly what is being tested, specifically the tests based on R_{im} and R_{if} test H_0: $\rho_{im}=\rho_{if}$, while the tests based on C_{im} and C_{if} test H_0: $(f_{im}/\rho_{im})=(f_{if}/\rho_{if})$. Thus these individual tests do not relate to separate survival or recovery rate parameters; hence they have no great value by themselves.

This test could be used to compare data sets from two areas (rather than two sexes). Assume there are r areas, and we have a recovery data set from each area. Let areas be indexed by a. The null hypothesis H_0 is that there are equal annual recovery and survival rate parameters in all areas. The alternative is that there are different annual parameters for at least two areas. Under this alternative, the MSS is

$$\begin{aligned}\mathcal{s}_a &= \{\mathcal{s}_1,\ldots,\mathcal{s}_r\} \\ &= \{R_{1a},\ldots,R_{ka},C_{1a},\ldots,C_{\ell-1,a}, a=1,\ldots,r\}\,.\end{aligned}$$

In this notation the second subscript denotes the area from where the basic data came. Under the null hypothesis H_0, the MSS is

$$\mathcal{s}_0 = \{(R_{11}+\cdots+R_{1r}),\ldots,(R_{k1}+\cdots+R_{kr}),(C_{11}+\cdots+C_{1r}),\ldots,(C_{\ell-1,1}+\cdots+C_{\ell-1,r})\}.$$

Using independence, and the known distribution under H_0 of each $\mathcal{s}_j$, $j=1,\ldots,r$, (a product of binomials) we can determine

$$P_{H_0}\{\mathcal{s}_a \mid \mathcal{s}_0\} = \prod_{i=1}^{k} \frac{\binom{N_{i1}}{R_{i1}}\cdots\binom{N_{ir}}{R_{ir}}}{\binom{N_{i1}+\cdots+N_{ir}}{R_{i1}+\cdots+R_{ir}}} \prod_{i=1}^{\ell-1} \frac{\binom{T_{i1}}{C_{i1}}\cdots\binom{T_{ir}}{C_{ir}}}{\binom{T_{i1}+\cdots+T_{ir}}{C_{i1}+\cdots+C_{ir}}}\,.$$

This distribution is a product of independent multiple hypergeometric distributions. The test of H_0 is based on the corresponding series of $k+\ell-1$ contingency tables, each r by 2 (as described in Section 5.1). The reader should be able to associate each of these r by 2 tables with the exact parametric function of survival and recovery rates which is being tested for equality (based on the distribution of the MSS).

Adult and Young Data

Next the basic theory is given for testing parameters for sex or area differences for recovery data sets from banding both adults and young. The H_1 model and the notation of Chapter 3, Sections 3.1 and 3.2 are used.

From Brownie (1973), an MSS under the H_1 model for a single data set with $s=0$ is (see also Sections 3.1 and 5.1 for definition of terms)

$$\mathcal{s} = \{R_{i\cdot}, Q_{i\cdot}, i=1,\ldots,k, Q_{ii}, i=1,\ldots,k-1, T_2, T_i+U_i, i=3,\ldots,k\}\,.$$

Additional statistics, computable from $\mathcal{s}$, and necessary in the tests are

$$\begin{aligned} W_i &= R_{\cdot i}+Q_{\cdot i}-Q_{ii} && ,i=1,\ldots,k, \\ X_i &= T_i+U_i-Q_{i\cdot} && ,i=1,\ldots,k\,. \end{aligned}$$

For the case of $s>0$, the MSS is equal to $\mathcal{s}$ given above, augmented by the additional terms Q_{kk}, $R_{\cdot k+j}+Q_{\cdot k+j}$, $j=1,\ldots,s$, and W_i, X_i, $i=k+1,\ldots,k+s$, where for $i>k$ we define $W_i=R_{\cdot i}+Q_{\cdot i}$ and $X_i=T_i+U_i$.

The MSS is distributed as the product of $4k+s-1$ independent binomial random variables if $s>0$, and as $4k-2$ independent binomial random variables if $s=0$;

$$\begin{aligned} &R_{i\cdot} \sim B(N_i,\rho_i) && ,i=1,\ldots,k, \\ &Q_{i\cdot} \sim B(M_i,\rho_i') && ,i=1,\ldots,k, \\ &Q_{ii} \text{ given } Q_{i\cdot} \sim B(Q_{i\cdot}, f_i'/\rho_i') && ,i=1,\ldots,k-1, \\ &W_i \text{ given } X_i \sim B(X_i, f_i/\rho_i) && ,i=1,\ldots,k+s-1\,, \end{aligned}$$

and if $s>0$ we also have

$$Q_{kk} \text{ given } Q_{k\cdot} \sim B(Q_{k\cdot}, f_k'/\rho_k')\,.$$

Now let there be r data sets, indexed by $a=1,\ldots,r$, where a becomes another index on all sample sizes, statistics, and parameters. Basically we just have r independent data sets and associated statistics. To derive the test of the null hypothesis H_0 that the underlying recovery and survival parameters $(f_i, f_i', S_i, S_i'$, see Section 3.2 for details) are the same for all r data sets we can use the probability distribution under H_0 of $\mathcal{s}_1,\ldots,\mathcal{s}_r$ (which is the MSS under the alternative) given $\mathcal{s}_0$. Here $\mathcal{s}_0$ is the MSS if H_0 is true; it is simply the "addition" (element by element) of the individual MSS, $\mathcal{s}_1$. For example, the first k elements of $\mathcal{s}_0$ are

$$R_{i\cdot 1}+R_{i\cdot 2}+\cdots+R_{i\cdot r}=\sum_{a=1}^{r}R_{i\cdot a} \qquad ,i=1,\ldots,k\,.$$

Under the null hypothesis H_0, the distribution of $\mathcal{s}_0$ is known. Finally, it is easy to determine

$$P_{H_0}\{\mathcal{s}_1,\ldots,\mathcal{s}_r \mid \mathcal{s}_0\}=\frac{\prod_{a=1}^{r} P_{H_0}\{\mathcal{s}_a\}}{P_{H_0}\{\mathcal{s}_0\}}\,.$$

The final result for $P_{H_0}\{\mathcal{s}_1,\ldots,\mathcal{s}_r \mid \mathcal{s}_0\}$ is the product of numerous terms, it can be expressed as

$$\prod_{i=1}^{k}\frac{\binom{N_{i1}}{R_{i\cdot 1}}\cdots\binom{N_{ir}}{R_{i\cdot r}}}{\binom{N_{i1}+\cdots+N_{ir}}{R_{i\cdot 1}+\cdots+R_{i\cdot r}}}\prod_{i=1}^{k}\frac{\binom{M_{i1}}{Q_{i\cdot 1}}\cdots\binom{M_{ir}}{Q_{i\cdot r}}}{\binom{M_{i1}+\cdots+M_{ir}}{Q_{i\cdot 1}+\cdots+Q_{i\cdot r}}}\times\prod_{i=1}^{\substack{k-1 \text{ if } s=0\\ k \text{ if } s>0}}\frac{\binom{Q_{i\cdot 1}}{Q_{ii1}}\cdots\binom{Q_{i\cdot r}}{Q_{iir}}}{\binom{Q_{i\cdot 1}+\cdots+Q_{i\cdot r}}{Q_{ii1}+\cdots+Q_{iir}}}\prod_{i=1}^{k+s-1}\frac{\binom{X_{i1}}{W_{i1}}\cdots\binom{X_{ir}}{W_{ir}}}{\binom{X_{i1}+\cdots+X_{ir}}{W_{i1}+\cdots+W_{ir}}}\,.$$

Each term gives rise to an r by 2 contingency table from which a chi-square test statistic can be computed. For example, the tables

$R_{i\cdot 1}$	$N_{i1}-R_{i\cdot 1}$
$R_{i\cdot 2}$	$N_{i2}-R_{i\cdot 2}$
.	.
.	.
.	.
$R_{i\cdot r}$	$N_{ir}-R_{i\cdot r}$

$,i=1,\ldots,k\,,$

derive from the first product term of the above distribution.

Reference to the original binomial distributions from whence a table derives will show what particular parametric function of survival and recovery rates is being tested for equality over the r data sets. For the above example, $\rho_{i1}=\rho_{i2}=\cdots=\rho_{ir}$ is being tested.