Adapting TP statistic to DIC model for Aids data

In DIC model the only choice for states j, l, and m, is $j = 1, l = 2$, and $m = 3$. The only null hypothesis for testing the Markov property against general alternative is tiis

$$H_0 : \alpha_{23}(t|X(s) = 1) = \alpha_{23}(t|X(s) = 2), t > s.$$

Proper and improper MI rectangles

We divide MI rectangles obtained from M observation rectangles into two subsets according to whether the time of 2 to 3 transition is finite or is right censored in state 2:

$$\{R_u \equiv (l_u, r_u] \times (w_u, z_u], z_u < \infty, 1 \leq u \leq U_1\}$$

and

$$R_u \equiv (l_u, r_u] \times (w_u, \infty], 1 \leq u \leq U_2\}.$$

We refer to an MI rectangle in the first subset as proper and in the second as improper. Similarly an observation rectangle (OR) is proper or improper depending on whether its time of 2 to 3 transition is finite or right censored in state 2

1. Reduction step

If $(l_u, r_u] \times (w_u, z_u]$ is the MI rectangle for a proper OR, we reduce it to the proper pair (r_u, z_u), thus assuming that 1 to 2 and 2 to 3 transitions occur at the end points of their respective intervals.

If $(l_u, r_u] \times (w_u, \infty]$ is the MI rectangle for an improper OR we reduce it to the improper pair (r_u, w_u) thus assuming that 1 to 2 transition occurs at the endpoint of its interval and keep the censoring time w_u in state 2 as is.

2. Specializing to Aids data. This dataset has $M = 96$ ORs of which $M_p = 27$ are proper and $M_{ip} = 69$ are improper rectangles. There are 25 MI rectangles, of which $m_p = 17$ are proper and $m_{ip} = 8$ are improper. The 17 proper MI rectangles are listed in Table 1 below and 8 improper MI rectangles are listed in Table 4 below. Among 17 proper MI rectangles, there are $V = 11$ unique times of the $2 \to 3$ transition. We denote these times by $t_1 < t_2 < ... < t_V$. They are listed in Table 3.

The Aids data are such that each proper OR rectangle contains only one of 17 proper MI rectangles. For each proper OR we have to know which proper MI it contains, see Table 2 below.

.In Aids data, each improper OR is of the form

$$(L_m, R_m] \times (23, \infty],$$

where $(L_m, R_m]$ is an interval in which $1 \to 2$ transition occurs after which a patient is right censored in state 2 by time 23, which is the end of study time. In particular if improper OR is of the form

$$(L_m, L_m + 1] \times (23, \infty] \tag{1}$$

it coincides with one of the improper MI rectangles. There are 30 improper ORs that satisfy this condition, see Table 5. There are 9 other improper ORs which contain only one improper MI rectangle, see Table 6. We refer to the improper OR which contains only one improper MI as type 1 improper OR. We refer to the remaining $69 - 39 = 30$ improper ORs which contain more than one.improper MI rectangle as type 2 improper ORs.

Computation of the TP statistic for all observations using $s = 8$. The logrank statistic is

$$\begin{aligned} \widehat{U}_8 &= \sum_{v=1}^{11} U_8(t_v) \qquad (2) \\ &= \sum_{v=1}^{11} \left\{ \sum_{i=1}^{27} I(X(8)=1) dN_i(t_v) - \left(\sum_{i=1}^{96} I(X(8)=1) Y_i(t_v) \right) \frac{\sum_{i=1}^{27} dN_i(t_v)}{\sum_{i=1}^{96} Y_i(t_v)} \right\}, \end{aligned}$$

where the integral with respect to t was replaced by the sum over $t_v's$. $N(t)$ is a counting process for $2 \to 3$ transition by time t, and $Y(t)$ is the number at risk for 2 to 3 transition at time $t-$. Here $s = 8$ was chosen so that for some proper reduced observations, $X(8) = 1$, and for the remaining reduced observations, $X(8) = 2$. Otherwise, (2) would be equal to zero.

For a given t_v

$$\sum_{i=1}^{27} I(X(8)=1) dN_i(t_v) = \sum_{i=1}^{27} I(8 \le r_i) dN_i(t_v), \tag{3}$$

the number of subjects who were in state 1 at time 8 and made $2 \to 3$ transition at time t_v, while

$$\sum_{i=1}^{27} dN_i(t_v) \tag{4}$$

is the total number of subjects who made $2 \to 3$ transition at time t_v. Next

$$\sum_{i=1}^{96} Y_i(t_v) = \sum_{i=1}^{27} I(r_i < t_v \le z_i) + \sum_{i=28}^{66} I(r_i < t_v)$$

$$+ \sum_{i=67}^{96} [I(R_i < t_v) + I(R_i \ge t_v) \frac{\sum_{k=1}^{K_i} I(r_{i,k} < t_v) P(MI_{i,k})}{P(OR_i)}]$$

is the total number of patients who are at risk for $2 \to 3$ transition at time t_v. The first part of above expression is the contribution from 27 proper ORs to the risk set at time t_v, the second is the contribution of 39 type 1 improper ORs, and the last one is the contribution of type 2 improper ORs, where K_i is the number of improper MI rectangles contained in type 2 improper OR_i and $r_{i,k}$ is the right endpoint of the k'th improper MI rectangle belonging to improper OR_i.

Finally, the number of patients at risk of the $2 \to 3$ transition at t_v and who were in state 1 at time $s = 8$.is

$$\sum_{i=1}^{96} I(8 \le r_i) Y_i(t_v) = \sum_{i=1}^{27} I(8 \le r_i < t_v \le z_i)$$

$$+ \sum_{i=28}^{66} [I(8 \le r_i < t_v)] \tag{5}$$

$$+ \sum_{i=67}^{96} [I(8 \le R_i < t_v) + I(R_i > t_v) \frac{\sum_{k=1}^{K_i} I(8 \le r_{i,k} < t_v) P(MI_{i,k})}{P(OR_i)}].$$

The four quantities (3)-(5) are computed using the information in Tables 1-10 in Section 4.

3 The standardized test statistic

Under H_0,the standardized $\widehat{U}_8$, namely

$$\widetilde{U}_8 = \frac{\widehat{U}_8}{\sqrt{V(\widehat{U}_8)}},$$

has asymptotically standard normal distribution, see Titman and Putter (2022), who also give the variance of $\widehat{U}_8$ as

$$V(\widehat{U}_8) = \sum_{v:t_v \geq 8} \frac{n_{v1}(n_v - n_{v1})}{(n_v)^2}, \tag{6}$$

where

$$n_v = \sum_{i=1}^{96} Y_i(t_v), n_{v1} = \sum_{i=1}^{96} I(8 \leq r_i) Y_i(t_v).$$

4. Data

Table 1 Proper MI rectangles with their probabilities

	Proper MI	p
1	$(5,7) \times (11,12)$	0.0208
2	$(5,7) \times (13,13)$	0.0208
3	$(1,7) \times (16,16)$	0.0104
4	$(3,7) \times (17,17)$	0.0104
5	$(7,9) \times (19,19)$	0.0104
6	$(9,10) \times (15,15)$	0.0156
7	$(10,11) \times (14,15)$	0.0156
8	$(10,11) \times (15,16)$	0.0312
9	$(9,11) \times (17,18)$	0.0312
10	$(10,12) \times (16,17)$	0.0104
11	$(9,12) \times (22,22)$	0.0104
12	$(10,12) \times (23,23)$	0.0104
13	$(12,13) \times (20,20)$	0.0104
14	$(13,14) \times (17,18)$	0.0312
15	$(13,14) \times (19,20)$	0.0104
16	$(13,14) \times (20,21)$	0.0208
17	$(14,15) \times (23,23)$	0.0104

Table 2 Proper OR(first column), with its proper MI rectangle(2nd column) and in reduced form (third column) There are 27 proper ORs listed below.

Proper OR	Proper MI	Reduced (r,z)
$1(5,7) \times (11,12)$	$(5,7) \times (11,12)$	$(7,12)$
$2(1,10) \times (11,12)$	$(5,7) \times (11,12)$	$(7,12)$
$3(5,8) \times (13,13)$	$(5,7) \times (13,13)$	$(7,13)$
$4(1,7) \times (12,13)$	$(5,7) \times (13,13)$	$(7,13)$
$5(9,13)\times(14,15)$	$(10,11) \times (14,15)$	$(11,15)$
$6(8,10) \times (15,15)$	$(9,10) \times (15,15)$	$(10,15)$
$7(10,11) \times (14,15)$	$(10,11) \times (14,15)$	$(11,15)$
$8(10,12) \times (15,16)$	$(10,11) \times (15,16)$	$(11,16)$
$9(10,14) \times (15,16)$	$(10,11) \times (15,16)$	$(11,16)$
$10(1,7) \times (16,16)$	$(1,7) \times (16,16)$	$(7,16)$
$11(10,11) \times (15,16)$	$(10,11) \times (15,16)$	$(11,16)$
$12(10,12) \times (16,17)$	$(10,12) \times (16,17)$	$(12,17)$
$13(3,7) \times (17,17)$	$(3,7) \times (17,17)$	$(7,17)$
$14(9,13) \times (17,18)$	$(9,11) \times (17,18)$	$(11,18)$
$15(13,14) \times (17,18)$	$(13,14) \times (17,18)$	$(14,18)$
$16(13,15) \times (17,18)$	$(13,14) \times (17,18)$	$(14,18)$
$17(9,12) \times (17,18)$	$(9,11) \times (17,18)$	$(11,18)$
$18(13,15) \times (17,18)$	$(13,14) \times (17,18)$	$(14,18)$
$19(9,11) \times (17,18)$	$(9,11) \times (17,18)$	$(11,18)$
$20(7,9) \times (19,19)$	$(7,9) \times (19,19)$	$(9,19)$
$21(12,13) \times (20,20)$	$(12,13) \times (20,20)$	$(13,20)$
$22(13,14) \times (19,20)$	$(13,14) \times (19,20)$	$(14,20)$
$23(1,15) \times (20,21)$	$(13,14) \times (20,21)$	$(14,21)$
$24\ (13,14) \times (20,21)$	$(13,14) \times (20,21)$	$(14,21)$
$25(9,12) \times (22,22)$	$(9,12) \times (22,22)$	$(12,22)$
$26(10,12) \times (23,23)$	$(10,12) \times (23,23)$	$(12,23)$
$27(14,15) \times (23,23)$	$(14,15) \times (23,23)$	$(15,23)$

Table 3 (For proper ORs) First column lists 11 unique times of $2 \rightarrow 3$ transition, second the number of patients that made the $2\rightarrow 3$ transition at t_v and were in state 1 at $s = 8$, third, the number of patients who are at risk for $2\rightarrow 3$ ar t_v, the last one gives the number of patients at risk at t_v who were in state 1 at s=8.

t_v	$\sum_{i=1}^{27} dN_i(t_v)$	$\sum_{i=1}^{27} I(8 \leq r_i)dN_i(t_v)$	$\sum_{i=1}^{27} Y_i(t_v)$	$\sum_{i=1}^{27} I(8 \leq r_i)Y_i(t_v)$
12	2	0	16	10
13	2	0	17	13
15	3	3	22	20
16	4	3	20	18
17	2	1	16	15
18	6	6	14	14
19	1	1	8	8
20	2	2	7	7
21	2	2	5	5
22	1	1	3	3
23	2	2	2	2

Note that multiplicities (entries in second column) add up to 27, which is the number of proper ORs. The entries in in Table 3 are obtained from Table 2 in the following way.

$$\sum_{i=1}^{27} dN_i(t_v) = \sum_{i=1}^{27} I(z_i = t_v),$$

$$\sum_{i=1}^{27} I(8 \leq r_i)dN_i(t_v) = \sum_{i=1}^{27} I(8 \leq r_i < z_i = t_v),$$

$$\sum_{i=1}^{27} Y_i(t_v) = \sum_{i=1}^{27} I(z_i \geq t_v),$$

$$\sum_{i=1}^{27} I(8 \leq r_i)Y_i(t_v) = \sum_{i=1}^{27} I(8 \leq r_i, z_i \geq t_v).$$

Table 4. Improper MI rectangles with their probabilities and in reduced form

	Improper MI rectangle (l,r)	probability	reduced $(r, 23)$
1	$(5, 7] \times (23, \infty)$	0.0437	$(7, 23)$
2	$(8, 9] \times (23, \infty)$	0.0235	$(9, 23)$
3	$(9, 10] \times (23, \infty)$	0.0639	$(10, 23)$
4	$(10, 11] \times (23, \infty)$	0.1649	$(11, 23)$
5	$(11, 12] \times (23, \infty)$	0.0379	$(12, 23)$
6	$(12, 13] \times (23, \infty)$	0.1682	$(13, 23)$
7	$(14, 15] \times (23, \infty)$	0.1548	$(15, 23)$
8	$(15, 16] \times (23, \infty)$	0.0619	$(16, 23)$

Table 5. Type 1 improper ORs which are of the form $OR_m = (L_m, L_m + 1] \times (23, \infty]$ with their probabilities and in reduced form.

Improper OR=Improper MI	multiplicity	reduced $(r, 23)$	p
$(9, 10) \times (23, \infty)$	2	$(10, 23)$	0.0639
$(10, 11) \times (23, \infty)$	7	$(11, 23)$	0.1649
$(11, 12) \times (23, \infty)$	1	$(12, 23)$	0.0379
$(12, 13) \times (23, \infty)$	7	$(13, 23)$	0.1682
$(14, 15) \times (23, \infty)$	8	$(15, 23)$	0.1548
$(15, 16) \times (23, \infty)$	5	$(16, 23)$	0.0619
total	30		

In "reduced" column, the entry $(10, 23)$ means that a patient made $1 \rightarrow 2$ transition at $r = 10$ and stayed in state 2 until the end of the study denoted by 23

Table 6. Type 1 improper ORs which are not of the form $OR_m = (L_m, L_m + 1] \times (23, \infty]$ with their probabilities and in reduced form.

Improper OR	Improper MI	multiplicity	reduced $(r, 23)$
$(1, 7) \times (23, \infty)$	$(5, 7) \times (23, \infty)$	1	$(7, 23)$
$(5, 7) \times (23, \infty)$	$(5, 7) \times (23, \infty)$	2	$(7, 23)$
$(7, 9) \times (23, \infty)$	$(8, 9) \times (23, \infty)$	1	$(9, 23)$
$(12, 14) \times (23, \infty)$	$(12, 13) \times (23, \infty)$	2	$(13, 23)$
$(13, 15) \times (23, \infty)$	$(14, 15) \times (23, \infty)$	3	$(15, 23)$
total		9	

We now compute the test statistic based on 66 ORs which contain only one MI rectangle assuming that $s = 8$..This statistic is

$$\widehat{U}_8 = \sum_{v=1}^{11} U_8(t_v),$$

where

$$U_8(t_v) = \sum_{i=1}^{27} I(8 \leq r_i) dN_i(t_v) - [\sum_{i=1}^{66} I(8 \leq r_i) Y_i(t_v)] \frac{\sum_{i=1}^{27} dN_i(t_v)}{\sum_{i=1}^{66} Y_i(t_v)}.$$

and is to be computed based on **Table 6a** below.

Table 6a The first two columns come from Table 3. Columns 3 and 4 use information from the same columns in Table 3 (first number). The second number is obtained. by inspection of Tables 5 and 6.

t_v	$\sum_{i=1}^{27} dN_i(t_v)$	$\sum_{i=1}^{27} I(8 \leq r_i) dN_i(t_v)$	$n_v = \sum_{i=1}^{66} Y_i(t_v)$	$n_{v1} = \sum_{i=1}^{66} I(8 \leq r_i) Y_i(t_v)$	$U_8(t_v)$
12	2	0	$16 + 13 = 29$	$10 + 10 = 20$	-1.37931
13	2	0	$17 + 14 = 31$	$13 + 11 = 24$	-1.54839
15	3	3	$22 + 23 = 45$	$20 + 20 = 40$	0.33333
16	4	3	$20 + 34 = 54$	$18 + 31 = 49$	-0.62963
17	2	1	$16 + 39 = 55$	$15 + 36 = 51$	-0.85454
18	6	6	$14 + 39 = 53$	$14 + 36 = 50$	0.33962
19	1	1	$8 + 39 = 47$	$8 + 36 = 44$	0.06383
20	2	2	$7 + 39 = 46$	$7 + 36 = 43$	0.13043
21	2	2	$5 + 39 = 44$	$5 + 36 = 41$	0.13636
22	1	1	$3 + 39 = 42$	$3 + 36 = 39$	0.07143
23	2	2	$2 + 39 = 41$	$2 + 36 = 38$	0.14634

Adding up the entries in the last column gives $\widehat{U}_8 = -3.19052$

To compute the variance of the test statistic $\widehat{U}_8$ we use

Table 6b The second and third columns are defined **in Table 6a.** The last column is obtained from expression for variance of the test statistic $\widehat{U}_8$ given in section 3 by (6).

t_v	n_V	n_{v1}	$n_{v1}(n_V - n_{v1})/n_V^2$
12	29	20	0.214031
13	31	24	0.174818
15	45	40	0.0987654
16	54	49	0.0840192
17	55	51	0.0674380
18	53	50	0.0533998
19	47	44	0.0597555
20	46	43	0.0609641
21	44	41	0.0635331
22	42	39	0.0663265
23	41	38	0.0678167

Adding up the entries in tha last column gives $V(\widehat{U}_8) = 1.01087$. Then the standardized test statistic is

$$\frac{\widehat{U}_8}{\sqrt{V(\widehat{U}_8}} = \frac{-3.19052}{\sqrt{1.01087}} = -3.1733,$$

which leads to the rejection of null hypothesis.