

# *Analysis of multivariate binomial data: case control or ascertainment sampling*

*Klaus Holst & Thomas Scheike*

*March 29, 2017*

---

## *Overview*

When looking at multivariate binomial data with the aim of learning about the dependence that is present, possibly after correcting for some covariates many models are available.

- Random-effects models logistic regression covered elsewhere (glmer in lme4).  
in the mets package you can fit the
- Pairwise odds ratio model
- Bivariate Probit model
  - With random effects
  - Special functionality for polygenic random effects modelling such as ACE, ADE ,AE and so forth.
- Additive gamma random effects model
  - Special functionality for polygenic random effects modelling such as ACE, ADE ,AE and so forth.

These last three models are all fitted in the mets package using composite likelihoods for pairs of data. The models can be fitted specifically based on specifying which pairs one wants to use for the composite score.

The models are described in further details in the binomial-twin vignette.

## *Case-Control Sampling*

Sometimes, pairs are recruited after a case-proband is selected. This proband, can be either a

- case: must be representative of cases  
or a
- control: must be representative of controls

First thinking about pairs, we estimate parameters using the conditional likelihood given sampling wick for a binary 2 x 2 table can be written as

$$\frac{P(i,j)}{P(j)}$$

the probability of seeing  $(i, j)$  for the pair, given that the proband was observed as  $(j)$ .

We note that if the marginal is known, or possibly estimated from the full cohort. Then we can estimate dependence parameters using just the terms  $P(i, j)$  for the pairs. We can thus ignore the special sampling for models with marginal specification. If the marginal can not be obtained from other sources we need to maximize the full-pairwise-likelihood in all parameters, that is both dependence and marginal parameters.

Similarly, one can select a whole family based on having selected a proband, that is selected a representative member of either cases or controls. In this case we fit the models by using composited likelihoods, considering all pairs that involves the probands. This will give some lacking efficiency compared to looking at the full likelihood of the family given the proband.

### *Ascertainment Sampling*

Similarly, in the setting of pairs we can select all pairs where there is at least one event of interest.

First thinking about pairs, we estimate parameters using the conditional likelihood given sampling with for a binary  $2 \times 2$  table can be written as

$$\frac{P(i, j)}{1 - P(0, 0)}$$

the probability of seeing  $(i, j)$  for the pair, given that it is sampled.

If the marginal can be estimated from a full sample we can then estimate the dependence parameter using the ascertainment likelihood.

Generally, when whole families are ascertained the computation of the true truncation probability can be hard to the fact that families are hard to define in the real world. Nevertheless, if a random sample of such family is at hand. We suggest to in these families take out all pairs that satisfies the ascertainment criterion. With a family, with given size  $n$  we have binary observations  $(Y_1, \dots, Y_n)$ . The family is sampled or a random sample of families such that

$$\sum_{i=1}^n Y_i \geq 1.$$

We let the conditional distribution given sampling, be denoted as

$$P^O(\cdot) = P(\cdot | \sum_{i=1}^n Y_i \geq 1)$$

Now, we note that all pairs within these family that satisfies that

$Y_i + Y_j \geq 1$ , will have distribution

$$\begin{aligned} P^O(Y_i = o_1, Y_j = o_2 | Y_i + Y_j \geq 1) &= \frac{P^O(o_1, o_2)}{P^O(Y_i + Y_j \geq 1)} \\ &= \frac{P(Y_i = o_1, Y_j = o_2, \sum_{i=1}^n Y_i \geq 1)}{P(Y_i + Y_j \geq 1, \sum_{i=1}^n Y_i \geq 1)} \\ &= \frac{P(Y_i = o_1, Y_j = o_2)}{P(Y_i + Y_j \geq 1)} = \frac{P(o_1, o_2)}{1 - P(0, 0)} \end{aligned}$$

since we only consider the probabilities where  $o_1 + o_2 \geq 1$ . Also here we could condition on covariates.

So considering these pairs, or a random sample of them should yield valid inference. When standard errors are computed we need to rely on GEE type arguments. An advantage of this is that the ascertainment probability is much easier to get for the pairs. Again using the pairwise structure will lead to loss of efficiency compared to using the full likelihood of the ascertained families.

### *The twin-stutter data*

We consider the twin-stutter where for pairs of twins that are either dizygotic or monozygotic we have recorded whether the twins are stuttering <sup>1</sup>

We here consider MZ and same sex DZ twins.

Looking at the data

---

```

1 library(mets)
2 data(twinstut)
3 twinstut$binstut <- 1*(twinstut$stutter=="yes")
4 twinstut <- subset(twinstut, zyg%in%c("mz", "dz"))
5 head(twinstut)

```

---

	tvparnr	zyg	stutter	sex	age	nr	binstut
1	2001005	mz	no	female	71	1	0
2	2001005	mz	no	female	71	2	0
3	2001006	dz	no	female	71	1	0
8	2001012	mz	no	female	71	1	0
9	2001012	mz	no	female	71	2	0
11	2001015	dz	no	male	71	1	0

- First, we select an ascertainment sample of the data, thus selecting a random sample of all ascertained pairs.
- Secondly, we select a case-control sample of this data to illustrate the use of the methods.

### *Ascertainment Sampling*

Selecting the ascertained pairs

---

```

1 library(mets)
2 data(twinstut)
3 twinstut$binstut <- 1*(twinstut$stutter=="yes")

```

---

```

4 twinstut <- subset(twinstut,zyg%in%c("mz","dz"))
5 dnumeric(twinstut) <- ~.
6 dfactor(twinstut,labels=c("DZ","MZ")) <- binzyg-zyg.n
7 ddrop(twinstut) <- ~"*.n"
8
9 twinstut <-
  ↳ dby(twinstut,binstut~tvparnr,stuttot=sum,nn=seq_along,n=length)
10 twina <- subset(twinstut,n==2 & stuttot>=1)

```

Selecting on the pairs where there is stuttering at taking a look at the tables of discordance and concordance for the twins.

```

1 twinda <- fast.reshape(twina,id="tvparnr")
2 twind <- fast.reshape(twinstut,id="tvparnr")
3 dtable(twind,"binst*"~I(stuttot1>=1))
4 dtable(twinda,~"binst*")

```

```
I(stuttot1 >= 1): FALSE
```

```

      binstut2    0
binstut1
0             6632

```

```
I(stuttot1 >= 1): TRUE
```

```

      binstut2    0    1
binstut1
0             0 289
1            281 111

```

```

      binstut2    0    1
binstut1
0             0 289
1            281 111

```

Now doing the analyses

### *Biprobbit model*

Looking at the full data for comparison. We estimate an unstructured probit model with different correlations for MZ and DZ twins.

```

1 b1 <- biprobbit(binstut~sex,~-1+binzyg,data=twinstut,id="tvparnr")
2 summary(b1)

```

	Estimate	Std.Err	Z	p-value
(Intercept)	-1.794821	0.023289	-77.066826	0.0000
sexmale	0.401430	0.030179	13.301756	0.0000
r:binzygDZ	0.132458	0.062516	2.118802	0.0341
r:binzygMZ	1.096915	0.073574	14.909085	0.0000

```

logLik: -4400.536  mean(score^2): 1.022e-06
n pairs
21288  7313

```

Contrast:

```

Dependence [binzygDZ]
Mean        [(Intercept)]

```

	Estimate	2.5%	97.5%
Rel.Recur.Risk	1.77662	0.92746	2.62577

OR	1.88752	1.09432	3.25566
Tetrachoric correlation	0.13169	0.00993	0.24960
Concordance	0.00235	0.00140	0.00393
Casewise Concordance	0.06456	0.03937	0.10413
Marginal	0.03634	0.03287	0.04016

Note, that the Casewise Concordance is a consistently estimated under complete ascertainment, i.e., when we consider a random sample of affected twins (at least one of the twins must have the event).

### *Odd-Ratio modelling*

First looking at the marginal model based on the full data we find the overall level of stuttering and also that males have a much higher stuttering risk.

---

```

1 margbin <- glm(binstut~factor(sex),data=twinstut,family=binomial())
2 summary(margbin)

```

---

Call:

```
glm(formula = binstut ~ factor(sex), family = binomial(), data = twinstut)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.4127	-0.4127	-0.2716	-0.2716	2.5763

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.28191	0.05000	-65.64	<2e-16 ***
factor(sex)male	0.86171	0.06211	13.87	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 9328.6 on 21287 degrees of freedom  
Residual deviance: 9124.7 on 21286 degrees of freedom  
AIC: 9128.7

Number of Fisher Scoring iterations: 6

First, fitting the OR model for MZ and DZ for the full data, we find that MZ have a much higher dependence than DZ twins.

---

```

1 theta.des <- model.matrix( ~-1+factor(zyg),data=twinstut)
2 bin <- binomial.twostage(margbin,data=twinstut,var.link=1,
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

---

Dependence parameter for Odds-Ratio (Plackett) model  
With log-link  
\$estimates

	theta	se
factor(zyg)dz	0.5238541	0.2400861
factor(zyg)mz	3.4930902	0.1865567

\$or

	Estimate	Std.Err	2.5%	97.5%	P-value
factor(zyg)dz	1.69	0.405	0.894	2.48	3.11e-05

```
factor(zyg)mz      32.89    6.135 20.862 44.91 8.31e-08
```

```
$type
```

```
[1] "plackett"
```

```
attr("class")
```

```
[1] "summary.mets.twostage"
```

Now, using the overall marginal we look at the adjusted likelihood and find very similar results on the ascertained sample. Note, that the marginals are crucial for this analysis to give useful results.

---

```
1 theta.des <- model.matrix( ~-1+factor(zyg),data=twina)
2 bina <- binomial.twostage(margbin,data=twina,var.link=1,
3   clusters=twina$tvparnr,theta.des=theta.des,
4   pair.ascertained=1)
5 summary(bina)
```

---

```
Dependence parameter for Odds-Ratio (Plackett) model
```

```
With log-link
```

```
$estimates
```

```
              theta      se
factor(zyg)dz 0.4874213 0.2472523
factor(zyg)mz 3.4753766 0.1985974
```

```
$or
```

```
      Estimate Std.Err   2.5% 97.5% P-value
factor(zyg)dz    1.63   0.403  0.839  2.42 5.24e-05
factor(zyg)mz   32.31   6.417 19.734 44.89 4.77e-07
```

```
$type
```

```
[1] "plackett"
```

```
attr("class")
```

```
[1] "summary.mets.twostage"
```

### Additive gamma modelling

First, again for comparison fitting the full data for the AE model. We get the size of the genetic variance in this model.

---

```
1 out <-
  ↳ twin.polygen.design(twinstut,id="tvparnr",zygname="zyg",zyg="dz",type="ae")
2 bintwin <- binomial.twostage(margbin,data=twinstut,
3   clusters=twinstut$tvparnr,detail=0,theta=c(0.1)/1,var.link=0,
4   random.design=out$des.rv,theta.des=out$pardes)
5 summary(bintwin)
```

---

```
Dependence parameter for Clayton-Oakes model
```

```
Variance of Gamma distributed random effects
```

```
$estimates
```

```
              theta      se
dependence1 0.9094847 0.09536268
```

```
$type
```

```
[1] "clayton.oakes"
```

```
$h
```

```
      Estimate Std.Err 2.5% 97.5% P-value
dependence1      1      0    1    1      0
```

```
$vare
```

NULL

\$vartot

	Estimate	Std.Err	2.5%	97.5%	P-value
p1	0.909	0.0954	0.723	1.1	1.47e-21

attr(,"class")

[1] "summary.mets.twostage"

We first here take a look at the marginal model for the ascertained sample, and note as expected that this sample gives highly biased estimates for the marginal model.

---

```

1 outa <-
  ↳ twin.polygen.design(twina,id="tvparnr",zygname="zyg",zyg="dz",type="ae")
2 marga <- glm(binstut~sex,data=twina,family=binomial())
3 summary(marga)

```

---

Call:

glm(formula = binstut ~ sex, family = binomial(), data = twina)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.334	-1.298	1.028	1.028	1.061

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	0.27895	0.08739	3.192	0.00141 **
sexmale	0.08242	0.11237	0.733	0.46328

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1851.8 on 1361 degrees of freedom

Residual deviance: 1851.2 on 1360 degrees of freedom

AIC: 1855.2

Number of Fisher Scoring iterations: 4

Now, using the overall marginal model we look at the adjusted likelihood and find very similar results on the ascertained sample. Note, that the marginals are crucial for this analysis to give useful results.

---

```

1 abintwin1 <- binomial.twostage(margbin,data=twina,
2
3     ↳ clusters=twina$tvparnr,detail=0,theta=c(0.1)/1,var.link=0,
4     ↳ random.design=outa$des.rv,theta.des=outa$pardes,pair.ascertained=1)
5 summary(abintwin1)

```

---

Dependence parameter for Clayton-Oakes model

Variance of Gamma distributed random effects

\$estimates

	theta	se
dependence1	0.8920274	0.09732786

\$type

[1] "clayton.oakes"

\$h

	Estimate	Std.Err	2.5%	97.5%	P-value
--	----------	---------	------	-------	---------

```

dependence1      1      0      1      1      0

$zare
NULL

$var
  Estimate Std.Err  2.5% 97.5% P-value
p1      0.892  0.0973 0.701  1.08 4.95e-20

attr("class")
[1] "summary.mets.twostage"

```

In fact for this model we can also do a full-MLE fitting jointly the dependence parameters and the marginal model. This is based on the twostage option (twostage=0 is MLE). Here the starting value is given at the marginal model for the ascertained model. This gives quite similar results to the previous analyses with a genetic variance around 1.

---

```

1 aabintwin1 <- binomial.twostage(marga,data=twina,
2     clusters=twina$tvparnr,detail=0,theta=c(0.1)/1,var.link=0,
3     random.design=outa$des.rv,theta.des=outa$parides,pair.ascertained=1,twostage=0)
4 summary(aabintwin1)
5 coef(marga)
6 coef(margbin)

```

---

```

Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
      theta      se
dependence1 1.014398 0.1045593

$type
[1] "clayton.oakes"

$h
      Estimate Std.Err  2.5% 97.5% P-value
dependence1      1      0      1      1      0

$zare
NULL

$var
  Estimate Std.Err  2.5% 97.5% P-value
p1      1.01  0.105 0.809  1.22 2.97e-22

$marginal
      Coef.      SE Robust SE      z P-val lower2.5% upper97.5%
dependence2 -4.350 0.08560  0.08560 -50.8    0   -4.520   -4.180
dependence3  0.551 0.00606  0.00606  90.8    0    0.539    0.563

attr("class")
[1] "summary.mets.twostage"
(Intercept)      sexmale
 0.2789484    0.0824214
(Intercept) factor(sex)male
 -3.2819072    0.8617053

```

### Case Control Sampling

First, taking out all cases and one control for each case, we establish the pairs of these probands. This is based on keeping track of the



twin related to the proband. Here using some utility functions in the mets packages.

Then we write up the random design vectors and the parameter design for each pair using the kinship coefficient.

When specifying the pairs in the case-control setup the second column should be the probands.

---

```

1 twinstut <-
  ↳ dby(twinstut, binstut~tvparnr, stuttot=sum, nn=seq_along, n=length)
2 twinstut <- subset(twinstut, n==2)
3 twinstut <- dtransform(twinstut, nnrow=1:nrow(twinstut))
4 twinstut <- dby(twinstut, binstut~tvparnr, nnn=seq_along)
5 twinstut <- dby2(twinstut, nnrow~tvparnr, pairnr=rev)
6
7 cases <- which(twinstut$binstut==1)
8 controls <- sample(which(twinstut$binstut==0), 1217)
9 rowsca <- with(twinstut, nnrow[cases])
10 rowsco <- with(twinstut, nnrow[controls])
11 rpairs <- c(rowsca, rowsco)
12 cc.pairs <- cbind( with(twinstut, pairnr.nnrow[rpairs]), rpairs)
13
14 ids <- sort(unique(c(cc.pairs)))
15 ###
16 pairsids <- c(cc.pairs)
17 pair.new <- matrix(fast.approx(ids, pairsids), ncol=2)
18 head(pair.new)
19
20 dataaid <- dsort(twinstut[ids,], "tvparnr")
21 dataaid=dtransform(dataaid, kinship=0.5)
22 dataaid=dtransform(dataaid, kinship=1, binzyg=="MZ")
23 kinship <- dataaid$kinship[pair.new[,2]]
24
25 out <- make.pairwise.design(pair.new, kinship, type="ae")
26 names(out)
27 out$random.des[, , 1]
28 out$theta.des[, , 1]
```

---

```

      [,1] [,2]
[1,]    6    5
[2,]   22   21
[3,]   24   23
[4,]   44   43
[5,]   48   47
[6,]   52   51
[1] "random.design" "theta.des"      "ant.rvs"
      [,1] [,2] [,3]
[1,]    1    1    0
[2,]    1    0    1
[1] 0.5 0.5 0.5
```

Now doing the analyses, first with know marginals, that is marginals from the full data. For this analysis, since marginals do not contain dependence parameters we do not need to specify that this is case-control sampling. Having a correct is crucial for this to work, but this is certainly often possible in register based studies where a full cohort is also available.

---

```

1 cc <- binomial.twostage(margbin, data=dataaid,
2                       clusters=dataaid$tvparnr,
3                       pairs=pair.new,
4                       random.design=out$random.design,
5                       theta.des=out$theta.des,
```

```

6         pairs.rvs=out$ant.rvs,
7         case.control=0,twostage=1)
8 summary(cc)

```

---

```

Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
      theta      se
dependence1 0.8572081 0.09509316

$type
[1] "clayton.oakes"

$h
      Estimate Std.Err 2.5% 97.5% P-value
dependence1      1      0      1      1      0

$vare
NULL

$var tot
      Estimate Std.Err 2.5% 97.5% P-value
p1      0.857  0.0951 0.671  1.04 1.98e-19

attr("class")
[1] "summary.mets.twostage"

```

We now do the same analysis specifying the case-control sampling. This should result in the same dependence parameters as is also the case.

```

1 cc3 <- binomial.twostage(margbin,data=dataid,
2                           clusters=dataid$tvparnr,
3                           pairs=pair.new,
4                           random.design=out$random.design,
5                           theta.des=out$theta.des,
6                           pairs.rvs=out$ant.rvs,
7                           case.control=1,twostage=1)
8 summary(cc3)

```

---

```

Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
      theta      se
dependence1 0.8572081 0.09509316

$type
[1] "clayton.oakes"

$h
      Estimate Std.Err 2.5% 97.5% P-value
dependence1      1      0      1      1      0

$vare
NULL

$var tot
      Estimate Std.Err 2.5% 97.5% P-value
p1      0.857  0.0951 0.671  1.04 1.98e-19

attr("class")
[1] "summary.mets.twostage"

```

This model can also be fitted using a full likelihood of both dependence parameters and marginal parameters. Here there is

no need to have a correctly specified marginal. We here use the marginal fitting from the case-control data as starting values. Again we find a genetic variance around 1. The marginal parameters are also consistent with the results from the full analyses for the marginal parameters.

---

```

1 marga <- glm(binstut~sex,data=dataid,family=binomial())
2 cc3 <- binomial.twostage(marga,data=dataid,
3     clusters=dataid$tvparnr,
4     pairs=pair.new,
5     random.design=out$random.design,
6     theta.des=out$theta.des,
7     pairs.rvs=out$ant.rvs,
8     case.control=1,twostage=0)
9 summary(cc3)

```

---

```

Dependence parameter for Clayton-Oakes model
Variance of Gamma distributed random effects
$estimates
              theta              se
dependence1 0.8701439 0.09457095

$type
[1] "clayton.oakes"

$h
      Estimate Std.Err 2.5% 97.5% P-value
dependence1      1      0      1      1      0

$vare
NULL

$vartot
      Estimate Std.Err 2.5% 97.5% P-value
p1      0.87  0.0946 0.685  1.06 3.55e-20

$marginal
      Coef.      SE Robust SE      z P-val lower2.5% upper97.5%
dependence2 -3.60 0.00903  0.00903 -399.0  0      -3.62      -3.58
dependence3  1.05 0.01960  0.01960  53.4  0       1.01       1.09

attr(,"class")
[1] "summary.mets.twostage"

```

---

When probands are related, here we may choose both case and controls from the same twin-pair then we need to adjust standard errors by grouping together contribution from related probands. This can be done using the `se.cluster` option that specifies how to cluster in the computation of the standard errors. In this case, however, this will be same as the clusters since these also are identical across pairs.

### *Combining Case Control and Ascertainment Sampling*

When specifying such models based on the pairs, it is in fact possible to combine ascertained pairs with case-control sampling by specifying vectors as the `case.control=c(1,0,1,0)` and `pair.ascertained=c(0,1,0,1)` arguments. Here with two case-control pairs, and two ascertained pairs.