A comparative study between a multi-stage adaptive pool testing model without test errors and the non-adaptive model

Okoth Annette W. [1*], Nyongesa kennedy L. [2] and Kwach Bonface O. [3]

[1,2]*Department of Mathematics, Masinde Muliro University of Science and Technology, P.O.Box 190-50100, Kakamega-Kenya*
[3]*Department of Mathematics, Kibabii University, P.O.Box 1699-50200, Bungoma-Kenya*
[*]*Corresponding author E-mail: annokoth73@gmail.com:*

Abstract

Pool testing for presence or absence of a trait is less expensive, less time consuming and therefore more cost effective. This study presents a multi-stage adaptive pool testing estimator $\hat{p}_{en}$ of prevalence of a trait in the absence of test errors. Pool testing is more efficient, less expensive and less time consuming. An increase in the number of stages improves the efficiency of the estimator, hence construction of a multi-stage model. The study made use of the Maximum Likelihood Estimate (MLE) method and Martingale method to obtain the adaptive estimator and Cramer-Rao lower bound method to determine the variance of the constructed estimator. Matlab and R, statistical softwares were used for Monte-carlo simulation and verification of the model, then analysis and discussion of properties of the constructed estimator in comparison with the non-adaptive estimator in the literature of pool testing done alongside provision of the confidence interval of the estimator. Results have shown that as the number of stages increases, the efficiency of the multi-stage adaptive estimator in the absence of test errors also increases in comparison with the non-adaptive estimator in the absence of test errors. This makes the multi-stage adaptive estimator better than the corresponding non-adaptive estimator in the literature of pool testing.

Keywords: Pool testing, Adaptive estimator, Test errors, Confidence interval

1. Introduction

Prevalence of defective units in a large population from accurate diagnostic tests is a fundamental risk assessment and management factor [8]. Estimation of defective units one-by-one is inefficient and uneconomical, considering that in a given population only a few individuals may be defective. It is against this background that pool testing comes in handy because it is more effective, less time consuming and less expensive [4]. Pool testing occurs when units from a population are pooled and tested as a group for the presence or absence of a particular trait. It also reduces the Mean Squared Error (MSE) of the estimates, hence it is more efficient, as was established by Sobel and Ellashoff, [11]. There are two forms of pool testing namely

  (i) Non-adaptive pool testing scheme
 (ii) Adaptive pool testing scheme

1.1. Non-adaptive testing scheme

In this testing scheme, a large population is divided in to n groups which are then subjected to testing [4]. When tested, a group can either test positive or negative and the outcome of the test aids in constructing the non-adaptive model.

1.2. Adaptive testing scheme

In this scheme a population is divided in to n groups, which are partitioned depending on the number of stages to be considered. Predetermined parameters are used to partition the groups and the number of partitioning parameters depends on the number of stages [8]. Partitioned groups are then tested at various stages for the presence or absence of a trait and the results used to construct the adaptive model.

1.3. Introduction of the model

In this study we obtain a multi-stage adaptive estimator $\hat{p}_{en}$ of prevalence of a trait in the absence of test errors, using the maximum likelihood estimate (MLE) method and investigate its effeciency in comparison with the non-adaptive estimator. The adaptive testing scheme involves testing groups in stages and updating group sizes from one stage to the next, with the group size at a stage depending on the outcome of the test(s) at the preceding stage(s). That is testing n_1 groups each of size k_1 at stage one; n_2 groups each of size k_2 at stage two; n_3 groups each of size k_3 at stage three and so on; where k_3 depends on both k_1 and k_2 while k_2 depends on k_1. For a general adaptive scheme, at stage i n_i groups each of size k_i, where k_i depends on $k_{i-1}, k_{i-2}, k_{i-3}, ...k_1$ are constructed. The constructed groups are then subjected to testing, where a group yields either a positive or a negative result. The number of groups, n_i is determined before the experiment is carried out while $k'_i s$ are sequentially determined as the experiment progresses [9].
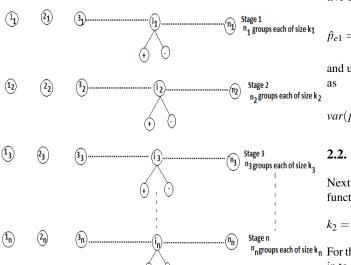
Figure 1: Multi-stage adaptive pool testing

2. Model Formulation and Analysis

We describe a multi-stage adaptive scheme with perfect tests as it is the backbone of this study and thereafter perform comparison analysis with other existing estimators, in the absence of test errors. For a multi-stage adaptive scheme, we set $n_1 = \lambda_1 n$, $n_2 = \lambda_2 n$, $n_3 = \lambda_3 n$, ..., $n_n = (1 - \lambda_1 n - \lambda_2 n - .... - \lambda_{n-1} n)$; where $\lambda_1, \lambda_2, ...$, λ_{n-1} are parameters used to partition the pools; k_2 depends on the outcome at stage 1, k_3 depends on the outcomes at stages 1 and 2 and k_n depends on the outcomes at stages $1, 2, 3, ..., n-1$. Each constructed group at each stage is then subjected to testing, yielding either a positive or negative result. This is shown in Figure 1 below: To construct this estimator, we consider the non-adaptive estimator in the absence of test errors, which according to [2] is given as

$$\hat{p}_e = 1 - \left(1 - \frac{X}{n}\right)^{\frac{1}{k}}. \tag{1}$$

and using Cramer-Rao lower bound method its variance is obtained as

$$var(\hat{p}_e) = \frac{1 - (1-p)^k}{nk^2(1-p)^{k-2}} \tag{2}$$

The non-adaptive estimator now enables us to explore the multi-stage adaptive estimator in the absence of test errors. This estimator is constructed through stages starting with stage one.

2.1. Adaptive Estimator at stage one

For the adaptive estimator at this stage, suppose X_1 out of n groups test positive on the test. Then X_1 has a Binomial distribution given by

$$X_1 \sim Binomial(\lambda n, \tau(p)|_{k=k_1}). \tag{3}$$

with its probability density function given by

$$f(x) = \binom{\lambda n}{x} \tau_1(p)^x (1 - \tau_1(p))^{\lambda n - x} \tag{4}$$

where,

$$\tau_1(p) = 1 - (1-p)^{k_1} \tag{5}$$

Using the Maximum Likelihood Estimate (MLE) method, the adaptive estimator at stage one is obtained as

$$\hat{p}_{e1} = 1 - \left(1 - \frac{X_1}{\lambda n}\right)^{\frac{1}{k_1}}. \tag{6}$$

and using Cramer-Rao lower bound method the variance is obtained as

$$var(\hat{p}_{e1}) = \frac{1 - (1-p)^{k_1}}{\lambda n k_1^2 (1-p)^{k_1 - 2}} \tag{7}$$

2.2. Adaptive Estimator at stage two

Next we determine the two-stage adaptive estimator which is a function of k_2, the group size at stage two and is determined by.

$$k_2 = argmin_l [Var(\hat{p}_1)]|_{p_1 = p}, \tag{8}$$

For the estimator at stage two, the total number of groups, n is divided in to λn groups each of size k_1 tested at stage one and $1 - \lambda n$ groups each of size k_2 tested at stage two. Suppose that X_2 groups out of the $(1 - \lambda)n$ groups tested at stage two are defective, then

$$X_2|X_1 \sim Binomial((1 - \lambda)n, \tau_{2|1}(p))$$

and the joint distribution of X_1 and X_2 is given as

$$f(X_1, X_2) = Binomial(\lambda n, \tau_1(p)|_{k=k_1}) * Binomial((1 - \lambda)n, \tau_{2|1}(p)). \tag{9}$$

Using the Maximum Likelihood Estimate (MLE) method, the adaptive estimator at stage two, $\hat{p}_{e2}$ of p is obtained as the solution to

$$\frac{k_1 * X_1}{1 - (1-p)^{k_1}} + \frac{k_2(X_1) * X_2}{1 - (1-p)^{k_2(X_1)}} = n(\lambda k_1 + (1 - \lambda)k_2(X_1)) \tag{10}$$

and using cramer-Rao lower bound method its variance is obtained as

$$Var(p_{e2}) = \frac{\tau_1(p)\tau_2(p)(1 - \tau_1(p))(1 - \tau_2(p))}{\tau_2(p)(1 - \tau_2(p))\lambda n k_1^2(1-p)^{2k_1 - 2} + \tau_1(p)(1 - \tau_1(p))(1 - \lambda)n k_2^2} \tag{11}$$

with τ_1 described in Equation (5) and τ_2 given as

$$\tau_{2|1}(p) = 1 - (1-p)^{k_2(X_1)}. \tag{12}$$

2.3. Adaptive Estimator at stage three

For the estimator at this stage, the total number of groups, n is divided in to $\lambda_1 n$ groups each of size k_1 tested at stage one, $\lambda_2 n$ groups each of size k_2 tested at stage two and $(1 - \lambda_1 - \lambda_2)n$ groups each of size k_3 tested at stage three, with k_3 described as

$$k_3 = argmin_l [Var(\hat{p}_2)]|_{p_2 = p_1}, \tag{13}$$

Suppose that X_3 groups out of the $(1 - \lambda_1 - \lambda_2)n$ groups tested at stage three are defective, then

$$X_3|X_2, X_1 \sim Binomial((1 - \lambda_1 - \lambda_2)n, \tau_{3|2,1}(p))$$

and the joint distribution of X_1, X_2 and X_3 is given as

$$\begin{aligned} f(X_1, X_2, X_3) &= Binomial(\lambda_1 n, \tau_1(p)|_{k=k_1}) \\ &* Binomial(\lambda_2 n, \tau_{2|1}(p)) \\ &* Binomial((1 - \lambda_1 - \lambda_2)n, \tau_{3|2,1}(p)). \end{aligned} \tag{14}$$

The adaptive estimator at stage three, $\hat{p}_{e3}$ of p is then obtained from Equation (14) using MLE method as the solution to

$$\frac{k_1 * X_1}{1 - (1-p)^{k_1}} + \frac{k_2(X_1) * X_2}{1 - (1-p)^{k_2(X_1)}} + \frac{k_3(X_2) * X_3}{1 - (1-p)^{k_3(X_2)}}$$
$$= n(\lambda_1 k_1 + \lambda_2 k_2(X_1) + (1 - \lambda_1 - \lambda_2)k_3(X_2)) \qquad (15)$$

and using cramer-Rao lower bound method its variance is obtained as

$$Var(p_{e3}) = \frac{\tau_1(p)\tau_2(p)\tau_3(p)(1 - \tau_1(p))(1 - \tau_2(p))(1 - \tau_3(p))}{E}, \qquad (16)$$

where E is described in the appendices. τ_1 and τ_2 are described by Equations (5), (12) while τ_3 is given by

$$\tau_{3|2,1}(p) = 1 - (1-p)^{k_3(X_2)} \qquad (17)$$

We can now consider the four-stage adaptive model in the absence of test errors and thereafter generalize to the multi-stage adaptive model

2.4. Adaptive Estimator at stage four

At this stage, the total number of groups, n is divided in to $\lambda_1 n$ groups each of size k_1 tested at stage one, $\lambda_2 n$ groups each of size k_2 tested at stage two, $\lambda_3 n$ groups each of size k_3 tested at stage three and $(1 - \lambda_1 - \lambda_2 - \lambda_3)n$ groups each of size k_4 tested at stage four. The group size, k_4 is described as

$$k_4 = argmin_l[Var(\hat{p}_3)]|_{p_3 = p_2}, \qquad (18)$$

. Now if X_4 groups out of the $(1 - \lambda_1 - \lambda_2 - \lambda_3)n$ groups tested at stage four are defective, then

$$X_4|X_3, X_2, X_1 \sim Binomial((1 - \lambda_1 - \lambda_2 - \lambda_3)n, \tau_{4|3,2,1}(p))$$

and the joint distribution of X_1, X_2, X_3 and X_4 is given as

$$\begin{aligned} f(X_1, X_2, X_3, X_4) &= Binomial(\lambda_1 n, \tau_1(p)|_{k=k_1}) \\ &* Binomial(\lambda_2 n, \tau_{2|1}(p)) \\ &* Binomial(\lambda_3 n, \tau_{3|2,1}(p)) \qquad (19) \\ &* Binomial((1 - \lambda_1 - \lambda_2 - \lambda_3)n, \tau_{4|3,2,1}(p)). \end{aligned}$$

The adaptive estimator at stage four, $\hat{p}_{e4}$ is then obtained from equation 18 as the solution to

$$\frac{k_1 * X_1}{1 - (1-p)^{k_1}} + \frac{k_2(X_1) * X_2}{1 - (1-p)^{k_2(X_1)}}$$
$$+ \frac{k_3(X_2) * X_3}{1 - (1-p)^{k_3(X_2)}} + \frac{k_4(X_3) * X_4}{1 - (1-p)^{k_4(X_3)}} \qquad (20)$$
$$= n(\lambda_1 k_1 + \lambda_2 k_2(X_1) + \lambda_3 k_3(X_2) + (1 - \lambda_1 - \lambda_2 - \lambda_3)k_4(X_3))$$

and using cramer-Rao lower bound method its variance is obtained as

$$Var(p_{e4}) = \frac{D}{F} \qquad (21)$$

where D and F are described in the appendices. τ_1, τ_2 and τ_3 are described by Equations (5), (12) and (17) while

$$\tau_{4|3,2,1}(p) = 1 - (1-p)^{k_4(X_3)} \qquad (22)$$

Using these ideas, we can now construct the multi-stage adaptive estimator, $\hat{p}_{en}$, determine its variance and do comparison analysis with other existing estimators.

2.5. Adaptive Estimator at stage n

Here the total number of groups, n is divided as described in sub-section 3.4 with the group size k_n determined as

$$k_n = argmin_l[Var(\hat{p}_{(n-1)})]|_{p_{(n-1)} = p_{(n-2)}}, \qquad (23)$$

Now if X_n groups out of the $(1 - \lambda_1 .... - \lambda_{n-1})n$ groups tested at stage n are defective, then following the steps followed in sub-section 3.4 above, the multi-stage adaptive estimator, $\hat{p}_{en}$ is obtained as the solution to

$$\frac{k_1 * X_1}{1 - (1-p)^{k_1}} + ... + \frac{k_n(X_{n-1}) * X_n}{1 - (1-p)^{k_n(X_{n-1})}}$$
$$= n(\lambda_1 k_1 + ... + (1 - \lambda_1 .... - \lambda_{n-1})k_n(X_{n-1})) \qquad (24)$$

while its variance is given as

$$Var(p_{en}) = \frac{\tau_1(p)....\tau_n(p)(1 - \tau_1(p))....(1 - \tau_n(p))}{G} \qquad (25)$$

G is given in the appendices.

3. Discussion of Results, Conclusion and Recommendations

In this section we discuss the results as provided by Table 4.1 and Figure 4.1. The highlights of the results would enable us make a detailed conclusion to this study. Here we compared our constructed estimators at stages two, three and four in the absence of test errors, that is $\hat{p}_{e2}$, $\hat{p}_{e3}$ and $\hat{p}_{e4}$ with the non-adaptive estimator in the absence of test errors. For the non-adaptive estimator in the absence of test errors, $var(\hat{p}_e)$ is given by Equation (2), [2] Asymptotic Relative Efficiency (ARE) values for stages two, three and four for the adaptive estimator in the absence of test errors are obtained by dividing Equation (2) by Equations (11), (16) and (21) respectively. That is,

$$\frac{Var(\hat{p}_e)}{Var(\hat{p}_{e2})}, \frac{Var(\hat{p}_e)}{Var(\hat{p}_{e3})} \text{ and } \frac{Var(\hat{p}_e)}{Var(\hat{p}_{e4})},$$

where $Var(\hat{p}_e)$, $Var(\hat{p}_{e2})$, $Var(\hat{p}_{e3})$ and $Var(\hat{p}_{e4})$ are given by Equations (2), (11), (16) and (21) respectively. On simplifying we obtain

$$ARE_{\hat{p}_{e2}} = \frac{U}{k^2(1-p)^{k-2}\tau_2\tau_1(1 - \tau_1)(1 - \tau_2)} \qquad (26)$$

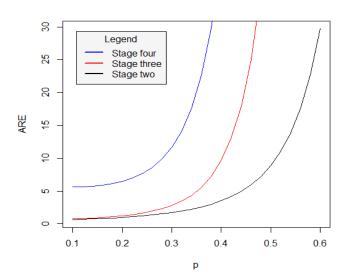$$ARE_{\hat{p}_{e3}} = \frac{V}{k^2(1-p)^{k-2}\tau_1\tau_2\tau_3(1 - \tau_1)(1 - \tau_2)(1 - \tau_3)} \qquad (27)$$

and

$$ARE_{\hat{p}_{e4}} = \frac{W}{k^2(1-p)^{k-2}\tau_1\tau_2\tau_3\tau_4(1 - \tau_1)(1 - \tau_2)(1 - \tau_3)(1 - \tau_4)} \qquad (28)$$

where U, V and W are desribed in the appendices. Using these Equations and R-Gui software Table 1 was generated.

p	$Stage2$	$Stage3$	$Stage4$
0.1	0.6472	0.7476	5.6716
0.2	0.9789	1.2274	6.4893
0.3	1.7039	2.7864	11.6500
0.4	3.5081	9.6128	40.3185
0.5	8.8854	52.0772	270.3464
0.6	29.6982	463.8524	3419.826

Table 1: ARE values of $\hat{p}_{e2}$, $\hat{p}_{e3}$ and $\hat{p}_{e4}$ relative to $\hat{p}_e$.

Figure 2: ARE of $\hat{p}_{e2}$, $\hat{p}_{e3}$ and $\hat{p}_{e4}$ vs probability, p

Table 1 provides generated ARE values of adaptive estimators at stages two, three and four, that is $\hat{p}_{e2}$, $\hat{p}_{e3}$ and $\hat{p}_{e4}$ relative to $\hat{p}_e$. From the table it is evident that these estimators register relatively high ARE values, except for $p = 0.1$ at stage two where the efficiency is slightly less than 1. This means that adaptive estimators in the absence of test errors are fairly more efficient than the non-adaptive estimator in the absence of test errors. It is also clear from the tables that ARE values increase with increase in the number of stages; the adaptive estimator at stage two having the lowest ARE values while the estimator at stage four has the highest ARE values. This is another important point to support the fact that the adaptive testing scheme gets better as the number of stages increases. This scenario is also depicted by Figure 2 below:

4. Conclusion and Recommendation

4.1. Conclusion

From our discussions, it is clear that as the number of stages increases the efficiency of the multi-stage adaptive estimators also increases. It is therefore evident that the multi-stage adaptive estimators out-perform the non-adaptive estimators. We therefore recommend the adoption of the multi-stage adaptive estimation scheme in estimating prevalence of a trait.

4.2. Recommendation For Further Work

This study focused on testing groups at various stages, with results at a stage depending on results at previous stages. However, a more inclusive scheme would consider retesting groups that test negative for example, with the view of checking quality to ensure that a positive group is not erroneously labeled as negative. This creates an opening for further research work.

Acknowledgement

References

[1] Bhattacharyya, G.K., Karandinos, M.G., and DeFoliart, G. R. (1979). *Point estimates and confidence intervals for infection rates using pooled organisms in epidemiologic studies.* American journal of epidemiology **109**,124-131.

[2] Brookmeyer, R. (1999). *Analysis of multistage pooling studies of biological specimens for estimating disease incidence and prevalence.* Biometrics **55**,608-612.

[3] Chiang, C. L. and Reeves, W. C. (1962). *Statistical estimation of virus infection rates in mosquito vector populations.* American journal on hygiene **75**, 377-391

[4] Dorfman R. (1943). *The detection of defective members of large populations.* Annals of mathematical statistics **14**, 436-440.

[5] Gastwirth, J.L., and Hammick,P.A. (1989). *Estimation of prevalence of a rare disease, preserving the anonymity of the subjects by group testing: Application to estimate the prevalence of AIDS antibodies in blood donors.* Journal of Statistical Planning and Inference **22**, 15-27.

[6] Mood, D. A., Graybill, G. A., Boes, D. C. (1974). Introduction to the theory of statistics. **27**, 72.

[7] Nyongesa, L. K. (2011). *Dual estimation of prevalence and disease incidence in pool testing strategy.* Communication in statistics theory and methods, **40**, 1-12

[8] Okoth, A. W. (2012). Two Stage Adaptive Pool Testing for estimating prevalence of a trait in the presence of test errors. Lambert Academic publishers, 18 and 27

[9] Oliver-Hughes J.M. and Swallow W.H. (1994), *A two-stage adaptive group testing procedure for estimating small proportions.* American statistical association **89**, 982-993.

[10] Richards, M. S. (1991). *Interpretation of the results of bacteriological testing of egg laying flocks for salmonella enteritidis.* Proceedings of the 6th international symposium on veterinary epidemiology and economics, 124-126, Ottawa, Canada.

[11] Sobel M., Ellashoff R.M. (1975).*Group testing with a new goal; estimation.* Biometrika **62**, 181-193.

[12] Swallow W.H. (1987).*Relative MSE and cost considerations in choosing group size for group testing to estimate infection rates and probability of disease infection.* Phytopathology **77**, 1376-1381.

Appendices

$$D = \tau_1(p)\tau_2(p)\tau_3(p)\tau_4(p)(1 - \tau_1(p))(1 - \tau_2(p))(1 - \tau_3(p))(1 - \tau_4(p))$$

$$
\begin{aligned}
E = & \ \tau_2(p)\tau_3(p)(1 - \tau_2(p))(1 - \tau_3(p))\lambda n k_1^2(1-p)^{2k_1-2} \\
+ & \ \tau_1(p)\tau_3(p)(1 - \tau_1(p))(1 - \tau_3(p))\lambda_2 n k_2(X_1)^2(1-p)^{2k_2(X_1)-2} \\
+ & \ \tau_1(p)\tau_2(p)(1 - \tau_1(p))(1 - \tau_2(p))(1 - \lambda_1 - \lambda_2)n k_3(X_2)^2(1-p)^{2k_3(X_2)}
\end{aligned}
$$

$$
\begin{aligned}
F = & \ \tau_2(p)\tau_3(p)\tau_4(p)(1 - \tau_2(p))(1 - \tau_3(p))(1 - \tau_4(p))\lambda n k_1^2(1-p)^{2k_1-2} \\
+ & \ \tau_1(p)\tau_3(p)\tau_4(p)(1 - \tau_1(p))(1 - \tau_3(p))(1 - \tau_4(p))\lambda_2 n k_2(X_1)^2(1-p) \\
+ & \ \tau_1(p)\tau_2(p)\tau_4(p)(1 - \tau_1(p))(1 - \tau_2(p))(1 - \tau_4(p))\lambda_3 n k_3(X_2)^2(1-p) \\
+ & \ \tau_1(p)\tau_2(p)\tau_3(p)(1 - \tau_1(p))(1 - \tau_2(p))(1 - \tau_3(p))(1 - \lambda_1 - \lambda_2 - \lambda_3)nk_4
\end{aligned}
$$

$$
\begin{aligned}
G = & \ \tau_2(p)....\tau_n(p)(1 - \tau_2(p))....(1 - \tau_n(p))\lambda n k_1^2(1-p)^{2k_1-2} \\
+ & \ ....\tau_1(p)....\tau_{n-1}(p)(1 - \tau_1(p))....(1 - \tau_{n-1}(p))(1 - \lambda_1.... - \lambda_{n-1})nk_n
\end{aligned}
$$

$$
\begin{aligned}
U = & \ 1 - (1-p)^k \Big[\tau_2(1-\tau_2)\lambda k_1^2(1-p)^{2k_1-2} \\
+ & \ \tau_1(1-\tau_1)(1-\lambda)k_2^2(1-p)^{2k_2-2} \Big]
\end{aligned}
$$

$$
\begin{aligned}
V = & \ 1 - (1-p)^k \Big[\tau_2\tau_3(1-\tau_2)(1-\tau_3)\lambda k_1^2(1-p)^{2k_1-2} \\
+ & \ \tau_1\tau_3(1-\tau_1)(1-\tau_3)\lambda_2 k_2^2(1-p)^{2k_2-2} \\
+ & \ \tau_1\tau_2(1-\tau_1)(1-\tau_2)(1-\lambda_1-\lambda_2)k_3^2(1-p)^{2k_3-2} \Big]
\end{aligned}
$$

$$
\begin{aligned}
W \;=\;\; & 1-(1-p)^k\Bigg[\tau_2\tau_3\tau_4(1-\tau_2)(1-\tau_3)(1-\tau_4)\lambda {k_1}^2(1-p)^{2k_1-2} \\[4pt]
+\;\; & \tau_1\tau_3\tau_4(1-\tau_1)(1-\tau_3)(1-\tau_4)\lambda_2{k_2}^2(1-p)^{2k_2-2} \\[2pt]
+\;\; & \tau_1\tau_2\tau_4(1-\tau_1)(1-\tau_2)(1-\tau_4)\lambda_3{k_3}^2(1-p)^{2k_3-2} \\[2pt]
+\;\; & \tau_1\tau_2\tau_3(1-\tau_1)(1-\tau_2)(1-\tau_3)(1-\lambda_1-\lambda_2-\lambda_3){k_4}^2(1-p)^{2k_4-2}\Bigg]
\end{aligned}
$$