

# Estimation of Aggregate Losses of Secondary Cancer Using PH-OPPL and PH-TPPL Distributions

Cynthia Mwende<sup>1\*</sup>, Patrick Weke<sup>2</sup>, Davis Bundi<sup>2</sup>, Joseph Ottieno<sup>2</sup>

<sup>1</sup>Department of Mathematics and Actuarial Science, Kisii University, Kisii, Kenya

<sup>2</sup>School of Mathematics, University of Nairobi, Nairobi, Kenya

Email: \*cynthiamwende@kisiuniversity.ac.ke, pweke@uonbi.ac.ke, dbundi@uonbi.ac.ke, sedimarvin254@gmail.com

**How to cite this paper:** Mwende, C., Weke, P., Bundi, D. and Ottieno, J. (2021) Estimation of Aggregate Losses of Secondary Cancer Using PH-OPPL and PH-TPPL Distributions. *Open Journal of Statistics*, 11, 838-853. <https://doi.org/10.4236/ojs.2021.115049>

**Received:** August 10, 2021

**Accepted:** October 16, 2021

**Published:** October 19, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Kenyan insurance firms have introduced insurance policies of chronic illnesses like cancer; however, they have faced a huge challenge in the pricing of these policies as cancer can transit into different stages, which consequently leads to variation in the cost of treatment. This has made the estimation of aggregate losses of diseases which have multiple stages of transitions such as cancer, an area of interest of many insurance firms. Mixture phase type distributions can be used to solve this setback as they can in-cooperate the transition in the estimation of claim frequency while also in-cooperating the heterogeneity aspect of claim data. In this paper, we estimate the aggregate losses of secondary cancer cases in Kenya using mixture phase type Poisson Lindley distributions. Phase type (PH) distributions for one and two parameter Poisson Lindley are developed as well their compound distributions. The matrix parameters of the PH distributions are estimated using continuous Chapman Kolmogorov equations as the disease process of cancer is continuous while severity is modeled using Pareto, Generalized Pareto and Weibull distributions. This study shows that aggregate losses for Kenyan data are best estimated using PH-OPPL-Weibull model in the case of PH-OPPL distribution models and PH-TPPL-Generalized Pareto model in the case of PH-TPPL distribution models. Comparing the two best models, PH-OPPL-Weibull model provided the best fit for secondary cancer cases in Kenya. This model is also recommended for different diseases which are dynamic in nature like cancer.

---

## Keywords

PH One Parameter Poisson Lindley, PH Two Parameter Poisson Lindley, PH Three Parameter Poisson Lindley, Discrete Fourier Transform, Discretization

---

## 1. Introduction

Aggregate losses are estimated by in-cooperating both claim frequency and claim severity distributions. Pavel (2010) [1] reviewed methods used to calculate distributions of aggregate losses. Robertson (1992) [2] applied Discrete Fourier Transform in estimation of aggregate losses from frequency and severity distributions. Rono *et al.* (2020) [3] developed compound distribution to model extreme natural disasters in Kenya. Mohamed *et al.* (2010) [4] introduced use of simulation approach in estimation of aggregate losses which can be employed when frequency and severity distribution cannot be combined to derive a compound distribution. Aggregate loss distributions are based on collective risk model expressed as:

$$S_N = \sum_{i=1}^N X_i \quad (1)$$

where:

$X_i$  is the severity distribution and  $N$  is the claim count distribution. The distribution of  $N$  in this paper is considered to follow mixed PH Poisson distributions.

Phase type distributions are constructed, when mixture distributions are convoluted resulting to an interrelated Poisson process occurring in phases. Phase type distributions were introduced way back by Erlang (1909) [5] and it has been advanced by Marcel F. Neuts (1981) [6] and Assussen (2003) [7] among others. Mogens Bladt (2005) [8] introduced phase type distributions in risk theory while O'cinneide (2017) [9] highlighted on Phase type distributions as well as their invariant polytopes. Wu *et al.* (2010) [10] developed phase type distributions when frequency distributions followed Panjer class  $(a, b, 0)$  while Kok *et al.* (2010) [11] used phase type distributions of Panjer class  $(a, b, 1)$  to model claim frequency.

Markov chains were introduced by Andrei Markov (1856-1922). Nurul *et al.* (2019) [12] proposed a simple forecasting model of predicting the future air quality using Markov chains which in-cooperated the Markov chains as an operator of evaluating pollution distribution in the long run. Yajuan *et al.* (2018) [13] used Markov chains to model demand for stations in Bike sharing systems. In this study, the concept of Markov chains is used to determine the matrices of the phase type distributions used in modeling claim frequency.

Frequency data is used to model occurrences in different areas such as engineering, insurance, biology etc. Poisson distribution is often used to model count data; however, it is based on the assumption that variance to mean ratio is unity (equi-dispersion) which is not applicable to real data; hence, it is considered as

an inflexible model. Most real life data either experience over dispersion where variance exceeds the mean or under dispersion where the mean exceeds the variance which can be modeled using Poisson mixtures [14]. Poisson Lindley distributions are perfect examples of Poisson mixtures where characteristics of Poisson distribution follow some characteristics of Lindley distribution. One parameter Poisson Lindley which can model over dispersed data was introduced by Sankaran (1970) [15] while Shanker and Mishra (2014) [16] developed two parameter Poisson Lindley which further research has justified that it can model over dispersed data.

In the insurance sector, when calculating aggregate losses for chronic diseases which have various stages like cancer the claim frequency distributions considered do not in-cooperate the different stages of such diseases. In-cooperating phase type distributions solve this short coming of ordinary distributions. Further considering mixed phase type distributions improves modeling of claim frequency data as it considers the heterogeneity aspect of claim data. In this paper, we develop PH one parameter Poisson Lindley distribution and PH two parameter Poisson Lindley distributions where the mixing distribution follows PH Lindley distribution. The resulting PH distributions are used to model claim numbers of secondary cancers in Kenya. Section 1 has a brief introduction to Poisson distributions and Poisson Lindley distributions.

The structure of this paper is as follows: Section 2 will discuss construction of phase type distribution using PH Lindley distributions which will later be applied in modeling of the aggregate losses. Compound distributions from the frequency and severity distributions are developed in Section 3. Aggregate losses for the data are estimated using Discrete Fourier Transforms and the results discussed in Section 4 and Section 5 outlines the conclusions.

## 2. Proposed Phase Type Poisson Lindley Distributions

In this section we develop phase type distributions for one parameter Poisson Lindley and two parameter Poisson Lindley. Phase type Poisson Lindley distributions are derived when the mixing distribution follow phase type Lindley distribution.

### 2.1. Phase Type One Parameter Poisson Lindley Distribution

**Definition 1.** A random variable  $X$  is said to be a phase type one parameter Poisson Lindley distribution if it follows:

$$\begin{aligned} X | \lambda &\sim Po(\lambda) \\ \lambda | \Lambda &\sim PH - OPL(\Lambda) \end{aligned}$$

for  $\lambda > 0$  and  $\Lambda$  is  $m * m$  matrix.

**Theorem 1.** If  $X \sim PH - OPPL$  distribution then the probability distribution function of  $X$  is:

$$f(x; \Lambda) = \bar{\gamma} \frac{\Lambda^2}{(I + \Lambda)^{x+3}} \{(x+2)I + \Lambda\} \bar{1}^T \quad (2)$$

where  $\Lambda$  is  $M * M$  and  $I$  is an identity matrix.

**Proof:**

If  $X | \lambda \sim Po(\lambda)$  and  $\lambda | \Lambda \sim PH-OPL(\Lambda)$ , then the pdf of variable  $X$  is expressed as;

$$P(x) = \int_0^{\infty} Pr(x | \lambda) f(\lambda; \Lambda) d\lambda$$

where  $f(\lambda; \Lambda)$  is  $PH-OPL(\Lambda)$ .

$$\begin{aligned} P(x) &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} \frac{\Lambda^2}{I + \Lambda} (1 + \lambda) e^{-\Lambda \lambda} \lambda > 0, \Lambda = M * M \\ &= \frac{\Lambda^2}{I + \Lambda} \int_0^{\infty} \left[ \frac{\lambda^x}{x!} e^{-\lambda(I + \Lambda)} + \frac{\lambda^{x+1}}{x!} e^{-\lambda(I + \Lambda)} \right] \\ &= \bar{\gamma} \frac{\Lambda^2}{(I + \Lambda)^{x+3}} \{(x + 2)I + \Lambda\} \bar{1}^T \end{aligned} \quad (3)$$

**Properties of Phase Type One Parameter Poisson Lindley Distribution**

The  $r^{th}$  moments of PH-OPPL distribution is given by:

$$\begin{aligned} E(X^r) &= \int_0^{\infty} x^r f(x, \Lambda) dx \\ &= \frac{\Lambda^2}{I + \Lambda} \int_0^{\infty} x^r e^{-\Lambda x} (1 + x) dx \\ &= \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Gamma(x + 1)}{\Lambda^{x+1}} + \frac{\Gamma(x + 2)}{\Lambda^{x+2}} \right] \\ &= \bar{\gamma} \frac{x! [(x + 1)I + \Lambda]}{\Lambda + I} \bar{1}^T \end{aligned} \quad (4)$$

The expectation and variance of PH-OPPL distribution can be easily obtained from Equation (4) as:

**1) Expectation**

$$E(x) = \frac{1! [(1 + 1)I + \Lambda]}{\Lambda (\Lambda + I)} = \bar{\gamma} \frac{2I + \Lambda}{\Lambda (\Lambda + I)} \bar{1}^T \quad (5)$$

**2) Variance**

$$\begin{aligned} Var(x) &= \frac{2! [(2 + 1)I + \Lambda]}{\Lambda^2 (\Lambda + I)} - \left\{ \frac{2I + \Lambda}{\Lambda (\Lambda + I)} \right\}^2 \\ &= \bar{\gamma} \frac{2I + 4\Lambda + \Lambda^2}{(\Lambda + I)^2} + \frac{2I + \Lambda}{\Lambda (\Lambda + I)} \bar{1}^T \end{aligned} \quad (6)$$

The probability generating function of PH-OPPL distribution is given by:

$$\begin{aligned} G(s) &= \int_0^{\infty} e^{\lambda(1-s)} \frac{\Lambda^2}{I + \Lambda} (1 + \lambda) e^{-\Lambda \lambda} d\lambda \\ &= \frac{\Lambda^2}{I + \Lambda} \left[ \int_0^{\infty} \lambda e^{-\lambda(I + \Lambda - sI)} d\lambda + \int_0^{\infty} e^{-\lambda(I + \Lambda - sI)} d\lambda \right] \\ &= \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + (2 - s)I}{[\Lambda + (1 - s)I]^2} \right] \bar{1}^T \end{aligned} \quad (7)$$

The parameter  $\Lambda$  of PH-OPPL distribution is estimated using continuous Chapman-Kolmogorov equation.

### 2.2. Phase Type Two Parameter Poisson Lindley Distribution

**Definition 2.** A random variable  $X$  is said to be a phase type two parameter Poisson Lindley distribution if it follows:

$$X | \lambda \sim Po(\lambda)$$

$$\lambda | \Lambda, \alpha \sim PH-TPL(\Lambda, \alpha)$$

for  $\alpha > 0, \lambda > 0$  and  $\Lambda$  is  $M * M$  matrix.

**Theorem 2.** If  $X \sim PH-TPPL$  distribution then the probability density function of  $X$  is expressed as:

$$f(x; \Lambda, \alpha) = \bar{\gamma} \frac{\Lambda^2}{(I + \Lambda)^{x+2}} \left[ I + \frac{(\alpha + x)I}{\alpha\Lambda + I} \right] \bar{1}^T \tag{8}$$

where  $\alpha > 0, \Lambda$  is  $M * M$  and  $I$  is an identity matrix.

**Proof:**

If  $X | \lambda \sim Po(\lambda)$  and  $\lambda | \Lambda, \alpha \sim PH-TPL(\Lambda, \alpha)$ , then the pdf of variable  $X$  is given by;

$$P(x) = \int_0^\infty Pr(X = x | \lambda) f(\lambda; \Lambda, \alpha) d\lambda$$

where  $f(\lambda; \Lambda, \alpha)$  is  $PH-TPL(\Lambda, \alpha)$ .

$$\begin{aligned} P(x) &= \int_0^\infty \frac{e^{-\lambda} \lambda^x}{x!} \frac{\Lambda^2}{I + \Lambda\alpha} (\alpha + \lambda) e^{-\Lambda\lambda} d\lambda \quad \lambda > 0, \Lambda = M * M \\ &= \frac{\Lambda^2}{I + \Lambda\alpha} \int_0^\infty \frac{\alpha\lambda^x}{x!} e^{-\lambda(I+\Lambda)} d\lambda + \int_0^\infty \frac{\lambda^{x+1}}{x!} e^{-\lambda(I+\Lambda)} d\lambda \\ &= \frac{\Lambda^2}{\alpha\Lambda + I} \left[ \frac{\alpha\Gamma(x+1)I}{x!(I+\Lambda)^{x+1}} + \frac{\Gamma(x+2)}{x!(I+\Lambda)^{x+2}} \right] \\ &= \bar{\gamma} \frac{\Lambda^2}{(I + \Lambda)^{x+2}} \left[ I + \frac{(\alpha + x)I}{\alpha\Lambda + I} \right] \bar{1}^T \end{aligned} \tag{9}$$

### Properties of Phase Type Two Parameter Poisson Lindley Distribution

The  $r^{th}$  moments of PH-TPPL distribution is given by:

$$\begin{aligned} E(X^r) &= \int_0^\infty x^r f(x, \Lambda, \alpha) dx \\ &= \int_0^\infty \left[ \sum_{x=0}^\infty x^r \frac{e^{-\lambda} \lambda^x}{x!} \right] \frac{\Lambda^2}{I + \Lambda\alpha} (\alpha + \lambda) e^{-\Lambda\lambda} d\lambda \\ &= \frac{\Lambda^2}{I + \Lambda\alpha} \int_0^\infty \lambda^r \Lambda e^{-\Lambda\lambda} d\lambda + \int_0^\infty \lambda^{r+1} e^{-\Lambda\lambda} d\lambda \\ &= \frac{\Lambda^2}{I + \Lambda\alpha} \left[ \alpha \frac{\Gamma(r+1)I}{\Lambda^{r+1}} + \frac{\Gamma(r+2)}{\Lambda^{r+2}} \right] \\ &= \bar{\gamma} \frac{\Gamma(r+1)I}{\Lambda^r} \frac{\alpha\Lambda + (r+1)I}{\alpha\Lambda + I} \bar{1}^T \end{aligned} \tag{10}$$

The expectation and variance of PH-TPPL distribution can be easily obtained from Equation (10) as:

**1) Expectation**

$$\begin{aligned}
 E(x) &= \frac{\Lambda^2}{\alpha\Lambda + I} \int_0^\infty \lambda(\alpha + \lambda)e^{-\Lambda\lambda} d\lambda \\
 &= \bar{\gamma} \frac{2I + \Lambda\alpha}{\Lambda(\Lambda\alpha + I)} \bar{1}^T
 \end{aligned}
 \tag{11}$$

**2) Variance**

$$\begin{aligned}
 Var(x) &= E(x^2) - [E(x)]^2 \\
 E(x^2) &= \frac{\alpha\Lambda + 2I}{\Lambda(\alpha\Lambda + I)} + \frac{2(\alpha\Lambda + 3I)}{\Lambda^2(\alpha\Lambda + I)} \\
 &= \bar{\gamma} \frac{\alpha\Lambda + 2I}{\Lambda(\alpha\Lambda + I)} + \frac{2(\alpha\Lambda + 3I)}{\Lambda^2(\alpha\Lambda + I)} \left[ \frac{2I + \Lambda\alpha}{\Lambda(\Lambda\alpha + I)} \right]^2 \bar{1}^T
 \end{aligned}
 \tag{12}$$

The probability generating function of PH-TPPL distribution is given by:

$$\begin{aligned}
 G(s) &= \frac{\Lambda^2}{(\Lambda + I)^2} \sum_{x=0}^\infty \left[ \frac{sI}{\Lambda + I} \right]^x \\
 &\quad + \frac{\Lambda^2}{(\Lambda + I)^2(\alpha\Lambda + I)} \sum_0^\infty (\alpha + x) \left[ \frac{s}{\Lambda + I} \right]^x \\
 &= \bar{\gamma} \frac{\alpha\Lambda[\Lambda + (1-s)I] + \Lambda^2}{(\alpha\Lambda + I)[\Lambda + (1-s)I]^2} \bar{1}^T
 \end{aligned}
 \tag{13}$$

The value of  $\Lambda$  is known hence the value of  $\alpha$  can be obtained from Equation (11) if the value of  $E(x)$  is known.

**2.3. Shape of Probability Function of PH-OPPL and PH-TPPL Distributions**

Matrix  $\Lambda$  was determined using continuous Chapman-Kolmogorov equation for cancer data in Kenya and the values of  $\gamma$  is the stationary probabilities obtained using the formula  $\pi_k = \pi_0 \Lambda^k$ . The values of  $\Lambda$  for three state Markov model represents cancer patients who transit from Healthy-Leukemia-Dead states, four state Markov model represents patients who transit from Healthy-Liver-Colon-Dead states, five state Markov model represents Healthy-Stomach-Pharynx-Colon-Dead states and six state Markov model represents patients transiting from Healthy-Oesophagus-Stomach-Lung-Kidney-Dead states. The values of  $\Lambda$  for different states are:

$$\begin{bmatrix} 0.8783 & 0.1217 & 0 \\ 0 & 0.3938 & 0.6062 \\ 0 & 0 & 1.0000 \end{bmatrix} \begin{bmatrix} 0.7900 & 0.2100 & 0 & 0 \\ 0 & 0.2898 & 0.7102 & 0 \\ 0 & 0 & 0.8985 & 0.1015 \\ 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

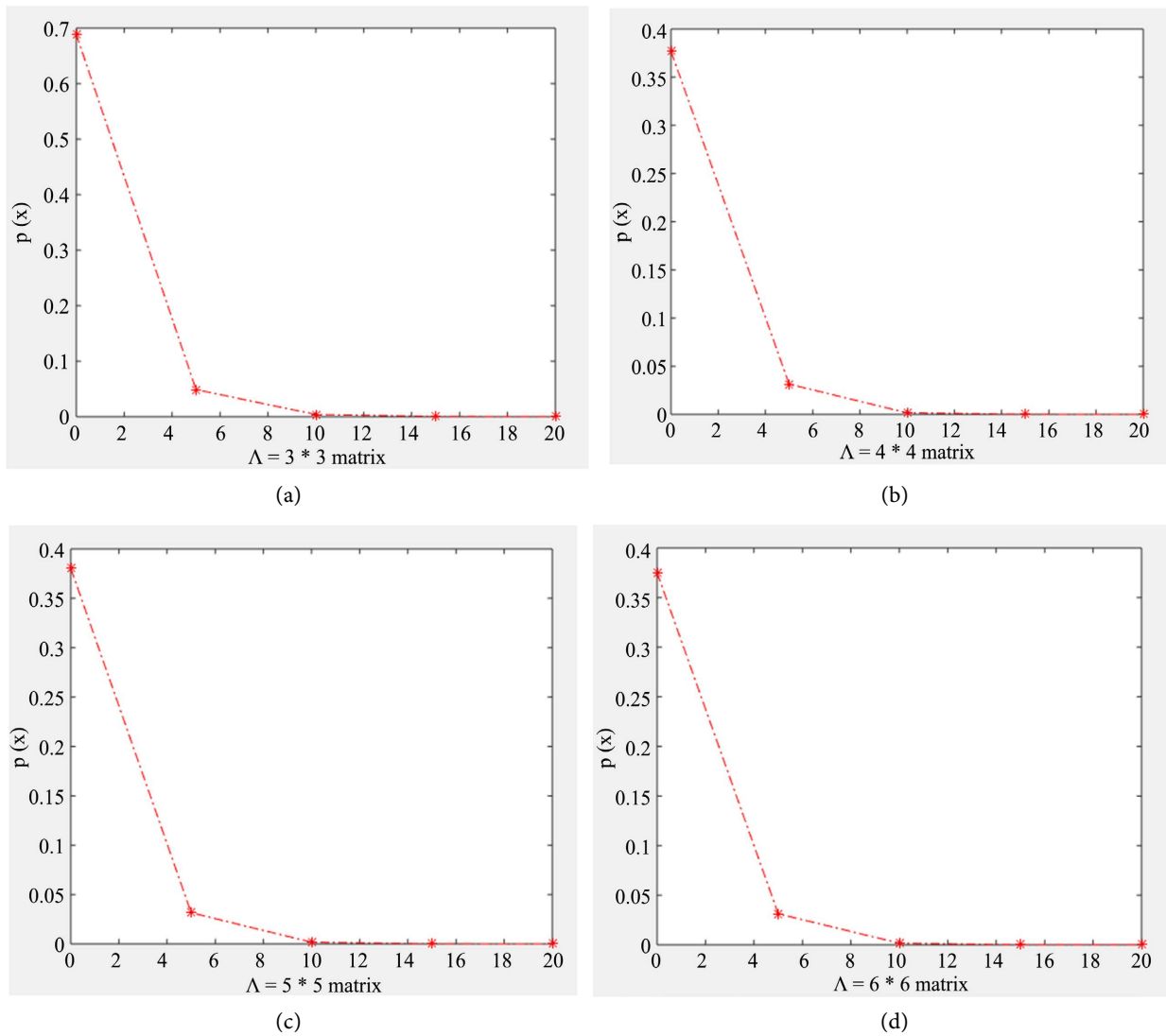
$$\begin{bmatrix} 0.8364 & 0.1636 & 0 & 0 & 0 \\ 0 & 0.3892 & 0.6108 & 0 & 0 \\ 0 & 0 & 0.6688 & 0.3312 & 0 \\ 0 & 0 & 0 & 0.5524 & 0.4476 \\ 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix} \begin{bmatrix} 0.4851 & 0.5149 & 0 & 0 & 0 & 0 \\ 0 & 0.1223 & 0.8777 & 0 & 0 & 0 \\ 0 & 0 & 0.1533 & 0.8467 & 0 & 0 \\ 0 & 0 & 0 & 0.4410 & 0.5590 & 0 \\ 0 & 0 & 0 & 0 & 0.8668 & 0.1332 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix}$$

The shape of probability function of phase type one parameter Poisson Lindley is expressed as:

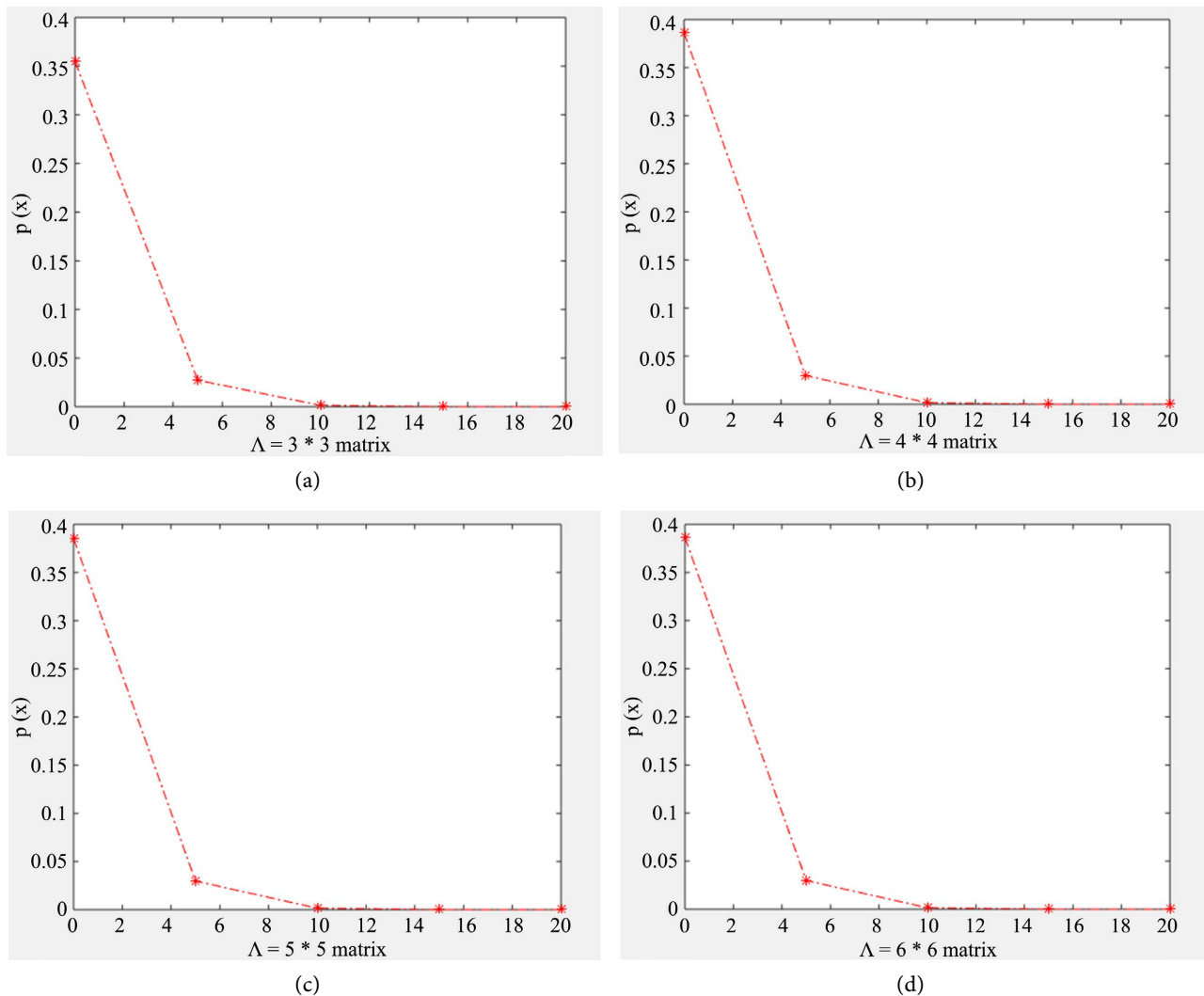
**Figure 1** shows that phase type one parameter Poisson Lindley is a long tailed distribution.

The shape of probability function of phase type two parameter Poisson Lindley is expressed as:

**Figure 2** shows that phase type two parameter Poisson Lindley is a long tailed distribution.



**Figure 1.** Pdf plots of PH-OPPL for different values of  $\Lambda$ .



**Figure 2.** Pdf plots of PH-TPPL for different values of  $\Lambda$ .

### 3. Compound Phase Type Distribution

Compound distribution in the actuarial field is the total losses in the group of insurance policies. In this section we develop compound phase type distributions (CPHD) which can be used to model secondary cancer cases.

**Definition 3.** Let  $N$  be a r.v with probability generating function  $F(S)$  and  $X_1, \dots, X_N$  be a set of iid random variable with a common probability generating function  $G(S)$  and is independent of  $N$ , then the probability generating function of the compound distribution is expressed as:

$$H(S) = F[G(S)] \quad (14)$$

Unlike ordinary compound distributions which do not consider transition phases of diseases, (CPHD) in-cooperates the transition states. Probability generating functions of compound distributions can be derived by convolution of probability generating function of two distributions as shown in Equation (14).

**Theorem 3 (Compound one parameter Poisson Lindley distribution).** If



the pgf of  $N \sim PH - OPPL(\Lambda)$  the compound pgf of  $N$  is:

$$H(S) = \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + (2 - L_x[G(S)])I}{[\Lambda + (1 - L_x[G(S)])I]^2} \right] \bar{1}^T \tag{15}$$

where  $L_x[G(S)]$  is the Laplace transform of the severity distribution as most continuous distributions their pgf is not available.

**Proof:**

$$\begin{aligned} H(S) &= F[G(S)] = F[L_x[G(S)]] \\ &= \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + (2 - L_x[G(S)])I}{[\Lambda + (1 - L_x[G(S)])I]^2} \right] \bar{1}^T \end{aligned} \tag{16}$$

**Theorem 4 (Compound two parameter Poisson Lindley distribution).** If the pgf of  $N \sim PH - TPPL(\Lambda)$  the compound pgf of  $N$  is:

$$H(S) = \bar{\gamma} \frac{\alpha\Lambda[\Lambda + (1 - L_x[G(S)])I] + \Lambda^2}{(\alpha\Lambda + I)[\Lambda + (1 - L_x[G(S)])I]^2} \bar{1}^T \tag{17}$$

where  $L_x[G(S)]$  is as defined in theorem (3).

**Proof:**

$$\begin{aligned} H(S) &= F[G(S)] = F[L_x[G(S)]] \\ &= \bar{\gamma} \frac{\alpha\Lambda[\Lambda + (1 - L_x[G(S)])I] + \Lambda^2}{(\alpha\Lambda + I)[\Lambda + (1 - L_x[G(S)])I]^2} \bar{1}^T \end{aligned} \tag{18}$$

The continuous distributions considered in this research are; Weibull, Pareto and Generalized Pareto distributions hence their Laplace transforms will be derived and replaced in Equations (16) and (18) to get the pgf of their compound distribution using PH-OPPL and PH-TPPL distributions respectively. The Laplace transform of Weibull, Pareto and Generalized Pareto are derived as:

**1) Weibull distribution**

$$\begin{aligned} L_x(S) &= E[e^{-sx}] \\ L_x G(S) &= \int_0^\infty e^{-sx} \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} dx \\ &= \frac{\beta}{\alpha} \int_0^\infty \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\frac{x}{\alpha} \left[s\alpha + \left(\frac{x}{\alpha}\right)^{\beta-1}\right]} dx \\ &= \frac{\beta}{\alpha} \frac{\Gamma\beta}{\left[s\alpha + \left(\frac{x}{\alpha}\right)^{\beta-1}\right]^\beta} \end{aligned} \tag{19}$$

**2) Pareto distribution**

$$L_x(S) = E[e^{-sx}]$$

$$\begin{aligned}
 L_x G(S) &= \alpha \beta^\alpha \int_0^\infty \frac{e^{-sx}}{(x+\beta)^{\alpha+1}} dx = \frac{\alpha}{\beta} \int_0^\infty e^{-\beta x} \sum_{k=0}^\infty \binom{-(\alpha+1)}{k} \left(\frac{x}{\beta}\right)^k dx \\
 &= \frac{\alpha}{\beta} \sum_{k=0}^\infty (-1)^k \frac{\Gamma(\alpha+k)}{k! \Gamma \alpha} \frac{k!}{\beta^{2k+1}} = \sum_{k=0}^\infty (-1)^k \frac{\alpha}{\Gamma \alpha} \frac{\Gamma(\alpha+k)}{\beta^{2k+2}}
 \end{aligned}
 \tag{20}$$

**3) Generalized Pareto distribution**

$$\begin{aligned}
 L_X(S) &= E[e^{-sX}] \\
 L_X G(S) &= \int_0^\infty e^{-sx} \frac{x^{\alpha-1}}{\beta(\alpha, \gamma)(x+\lambda)^{\alpha+\gamma}} dx \\
 &= \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \int_0^\infty x^\alpha e^{-sx} \sum_{k=0}^\infty \binom{\alpha+\gamma}{k} \frac{x^k}{\lambda^k} dx \\
 &= \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^\infty \frac{-\binom{\alpha+\gamma}{k}}{\lambda^k} \int_0^\infty x^{\gamma+k+1-1} e^{-sx} dx \\
 &= \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^\infty \frac{-\binom{\alpha+\gamma}{k} \Gamma \alpha + k}{\lambda^k s^{\alpha+k}}
 \end{aligned}
 \tag{21}$$

Replacing Equations (19), (20) and (21) in Equation (16) the pgf of the compound distributions of PH-one parameter Poisson Lindley with Weibull, Pareto and Generalized Pareto respectively are:

**1) Compound PH-OPPL-Weibull distribution**

$$H(S) = \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + \left( 2 - \frac{\beta}{\alpha} \frac{\Gamma \beta}{\left[ s\alpha + \left(\frac{x}{\alpha}\right)^{\beta-1} \right]^\beta} \right) I}{\left[ \Lambda + \left( 1 - \frac{\beta}{\alpha} \frac{\Gamma \beta}{\left[ s\alpha + \left(\frac{x}{\alpha}\right)^{\beta-1} \right]^\beta} \right) I \right]^2} \right] \bar{1}^T
 \tag{22}$$

**2) Compound PH-OPPL-Pareto distribution**

$$H(S) = \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + \left( 2 - \sum_{k=0}^\infty (-1)^k \frac{\alpha}{\Gamma \alpha} \frac{\Gamma(\alpha+k)}{\beta^{2k+2}} \right) I}{\left[ \Lambda + \left( 1 - \sum_{k=0}^\infty (-1)^k \frac{\alpha}{\Gamma \alpha} \frac{\Gamma(\alpha+k)}{\beta^{2k+2}} \right) I \right]^2} \right] \bar{1}^T
 \tag{23}$$

**3) Compound PH-OPPL-Generalized Pareto distribution**

$$H(S) = \bar{\gamma} \frac{\Lambda^2}{I + \Lambda} \left[ \frac{\Lambda + \left( 2 - \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^\infty \frac{-\binom{\alpha+\gamma}{k} \Gamma \alpha + k}{\lambda^k s^{\alpha+k}} \right) I}{\left[ \Lambda + \left( 1 - \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^\infty \frac{-\binom{\alpha+\gamma}{k} \Gamma \alpha + k}{\lambda^k s^{\alpha+k}} \right) I \right]^2} \right] \bar{1}^T
 \tag{24}$$

Replacing Equations (19), (20) and (21) in Equation (18) the pgf of the compound distributions of PH-two parameter Poisson Lindley with Weibull, Pareto and Generalized Pareto respectively are:

**1) Compound PH-TPPL-Weibull distribution**

$$H(S) = \bar{\gamma} \frac{\alpha\Lambda \left[ \Lambda + \left( 1 - \frac{\beta}{\alpha} \frac{\Gamma\beta}{\left[ s\alpha + \left( \frac{x}{\alpha} \right)^{\beta-1} \right]^\beta} \right) I \right] + \Lambda^2}{(\alpha\Lambda + I) \left[ \Lambda + \left( 1 - \frac{\beta}{\alpha} \frac{\Gamma\beta}{\left[ s\alpha + \left( \frac{x}{\alpha} \right)^{\beta-1} \right]^\beta} \right) I \right]^2} \bar{1}^T \tag{25}$$

**2) Compound PH-TPPL-Pareto distribution**

$$H(S) = \bar{\gamma} \frac{\alpha\Lambda \left[ \Lambda + \left( 1 - \sum_{k=0}^{\infty} (-1)^k \frac{\alpha}{\Gamma\alpha} \frac{\Gamma(\alpha+k)}{\beta^{2k+2}} \right) I \right] + \Lambda^2}{(\alpha\Lambda + I) \left[ \Lambda + \left( 1 - \sum_{k=0}^{\infty} (-1)^k \frac{\alpha}{\Gamma\alpha} \frac{\Gamma(\alpha+k)}{\beta^{2k+2}} \right) I \right]^2} \bar{1}^T \tag{26}$$

**3) Compound PH-TPPL-Generalized Pareto distribution**

$$H(S) = \bar{\gamma} \frac{\alpha\Lambda \left[ \Lambda + \left( 1 - \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^{\infty} \frac{-(\alpha+\gamma) \Gamma\alpha+k}{\lambda^k s^{\alpha+k}} \right) I \right] + \Lambda^2}{(\alpha\Lambda + I) \left[ \Lambda + \left( 1 - \frac{1}{\lambda^\gamma \beta(\alpha, \gamma)} \sum_{k=0}^{\infty} \frac{-(\alpha+\gamma) \Gamma\alpha+k}{\lambda^k s^{\alpha+k}} \right) I \right]^2} \bar{1}^T \tag{27}$$

**4. Data Analysis, Results and Discussions**

**4.1. Severity and Frequency Probabilities**

The cancer data considered in this research is obtained from a medical facility in Kenya. The cancer transitions states considered are Healthy-Leukemia-Dead states for 3 state model, Healthy-Liver-Colon-Dead states for four state model, Healthy-Stomach-Pharynx-Colon-Dead states for five state model and Healthy-Oesophagus-Stomach-Lung-Kidney-Dead states for six state models. The values of  $\Lambda$  for the data are obtained using continuous Chapman-Kolmogorov equations expressed as:

$$\begin{aligned} P_{ij}(\phi_A, \gamma_t + \Psi_d) &= \sum_{k=1}^n p_{ik}(\phi_A, \gamma_t) p_{kj}(\gamma_t, \gamma_t + \Psi_d) \\ \lim_{\Psi_d \rightarrow 0} \frac{P_{ij}(\phi_A, \gamma_t + \Psi_d) - P_{ij}(\phi_A, \gamma_t)}{\Psi_d} &= \lim_{\Psi_d \rightarrow 0} \frac{p_{ik}(\phi_A, \gamma_t) p_{kj}(\gamma_t, \gamma_t + \Psi_d) - p_{ij}(\phi_A, \gamma_t) [1 - p_{jj}(\gamma_t, \gamma_t + \Psi_d)]}{\Psi_d} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \gamma_t} P_{ij}(\phi_A, \gamma_t) &= \sum_{k=1}^n P_{ik}(\phi_A, \gamma_t) \mathfrak{S}_{kj} - P_{ij}(\phi_A, \gamma_t) \mathfrak{S}_j \\ P_{ij}(\phi_A, \gamma_t + \Psi_d) &= \sum_{k=1}^n P_{ik}(\phi_A, \gamma_t) P_{kj}(\gamma_t, \gamma_t + \Psi_d) \\ P_{ij}(\phi_A, \gamma_t) &= 1 - e^{-\mathfrak{S}_j(\phi_A)t} \end{aligned} \tag{28}$$

where:

$$\begin{aligned} \lim_{\Psi_d \rightarrow 0} \frac{P_{kj}(\gamma_t, \gamma_t + \Psi_d)}{\Psi_d} &= \mathfrak{S}_{kj} \\ \lim_{\Psi \rightarrow 0} \frac{1 - P_{ij}(\gamma_t, \gamma_t + \Psi_d)}{\Psi_d} &= \mathfrak{S}_j \end{aligned}$$

The values of  $\Lambda$  for three, four, five and six state using the data obtained were as shown in Section 2.3.

The severity distributions considered in this research are Weibull, Pareto and Generalized Pareto distributions. DFT requires severity probabilities to be discrete hence they will be discretized using method of mass rounding which is expressed as:

$$\begin{aligned} f_0 &= F_J\left(\frac{h}{2}\right) \\ f_x &= F_J\left(xh + \frac{h}{2}\right) - F_J\left(xh + \frac{h}{2}\right) \quad x = 1, 2, 3, \dots \\ f_m &= 1 - F_J\left(mh - \frac{h}{2}\right) \end{aligned}$$

The pdf of Wei-bull, Pareto and Generalized Pareto distributions respectively are expressed as;

$$\begin{aligned} f(x) &= \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}; \quad x > 0; a, b > 0 \\ f(x) &= \frac{\alpha \beta^\alpha}{x + \beta^{\alpha+1}} \\ f(x) &= \frac{x^{\alpha-1} \lambda^\gamma \Gamma(\alpha + \gamma)}{\Gamma \lambda \Gamma \alpha (x + \lambda)^{\alpha+\gamma}} \end{aligned}$$

The frequency and severity probabilities for secondary cancer cases are: (Table 1).

**Table 1.** Claim frequency and severity probabilities.

States	PH-OPPL	PH-TPPL	Weibull	Pareto	Gen Pareto
3 state	0.011	0.0078	0.1506	0.1099	0.0192
4 state	0.0082	0.0054	0.1266	0.0591	0.0091
5 state	0.0083	0.0055	0.1703	0.1904	0.0429
6 state	0.0.0083	0.0054	0.1775	0.2488	0.0730

## 4.2. Discrete Fourier Transform

There are different numerical methods used in estimation of aggregate losses such as; Monte Carlo, Panjer recursive model, Fourier transforms and Direct Numerical Integration. Panjer recursive model is applicable when the claim frequency distributions follow either Panjer class  $(a, b, 0)$  or class  $(a, b, 1)$ . In this study we will consider Discrete Fourier Transform (DFT) in estimation of the aggregate losses. Robertson (1992) applied Fourier transforms in computation of aggregate losses [2]. Pavel (2010) [1] reviewed these numerical methods and concluded that each method had its strength and weaknesses hence they should be chosen according to the study. DFT is mostly preferred as it is arguably said to be the most elegant and powerful technique in evaluation of aggregate loss probabilities when claim amount  $X_i$  is both discrete and continuous [17].

The algorithm of DFT of aggregate losses requires computation of DFT of frequency and DFT of severity separately.

**Definition 4 (Discrete Fourier Transform).** Let  $X_n$  be the severity or frequency distribution of the claim data. For any discrete function  $X_k$  the Discrete Fourier transform is the mapping;

$$X_k = \sum_{n=0}^{N-1} X(n) e^{-i2\pi kn/N} \quad k = 0, 1, 2, \dots, N-1 \quad (29)$$

Expression (29) is very complex to work with hence to reduce its complexity we apply Euler's formula and it becomes:

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} X(n) \cos\left(\frac{2\pi kn}{N}\right) + i \sin\left(\frac{2\pi kn}{N}\right) \\ X(k) &= \sum_{n=0}^{N-1} X(n) W_N^{kn} \end{aligned} \quad (30)$$

which is the DFT of the severity or frequency probabilities. The severity and frequency probabilities are of length 8 and hence the matrix  $W$  must be a primitive 8<sup>th</sup> root of unity therefore Equation (30) can be rewritten as:

$$X(k) = \sum_{n=0}^7 X(n) W_N^{kn} \quad (31)$$

The frequency or severity probabilities will be padded with equal number of zero's as its elements in order to perform no wrap convolution. The DFT algorithm is as follows:

- 1) Multiply the matrix  $W_N^{kn}$  with the frequency or severity probabilities to get the DFT of frequency or severity probabilities.
- 2) Compute DFT of DFT of frequency and severity by multiplying DFT of frequency probabilities with the DFT of the severity probabilities and consequently multiplying the resulting vector with the matrix  $W_N^{kn}$ .
- 3) Select the values without the complex  $i$  and divide each value by the number of elements in the vector of frequency or severity distribution and arrange the resulting probabilities in reverse except for the first probability.

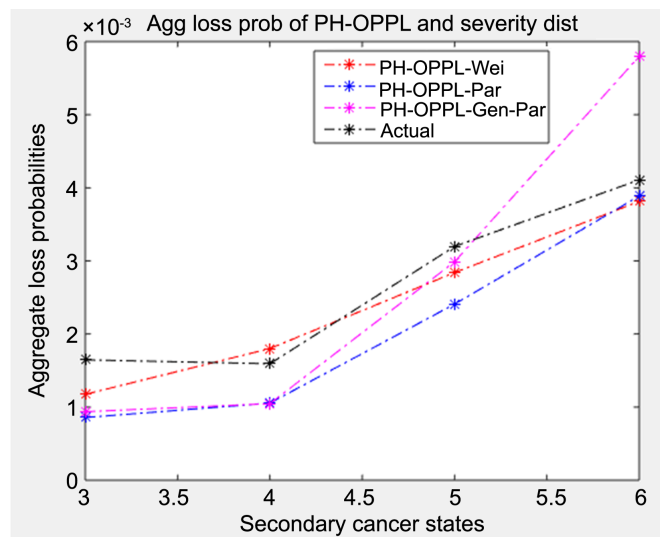
4) Values corresponding to original frequency and severity values are the aggregate loss probabilities.

The values of aggregate loss probabilities using DFT are:

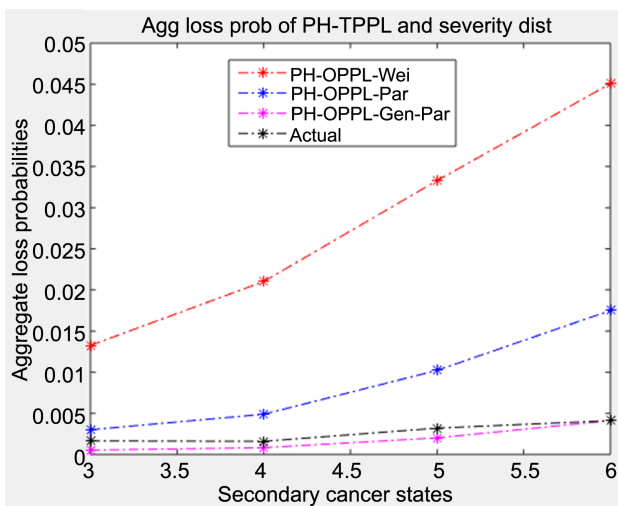
**Table 2.** Aggregate loss probabilities.

States	PH-OPPL/Wei	PH-OPPL/Par	PH-OPPL/Gen-par	PH-TPPL/Wei	PH-TPPL/Par	PH-TPPL/Gen-par
3 state	0.00117468	0.00085722	0.00093696	0.0132528	0.00300027	0.00052416
4 state	0.00180072	0.00105444	0.00104696	0.02102016	0.00488845	0.00082059
5 state	0.00284028	0.00240871	0.0029879	0.00332912	0.0102232313	0.00201259
5 state	0.00381366	0.00388731	0.00579889	0.04507728	0.01749645	0.00411375

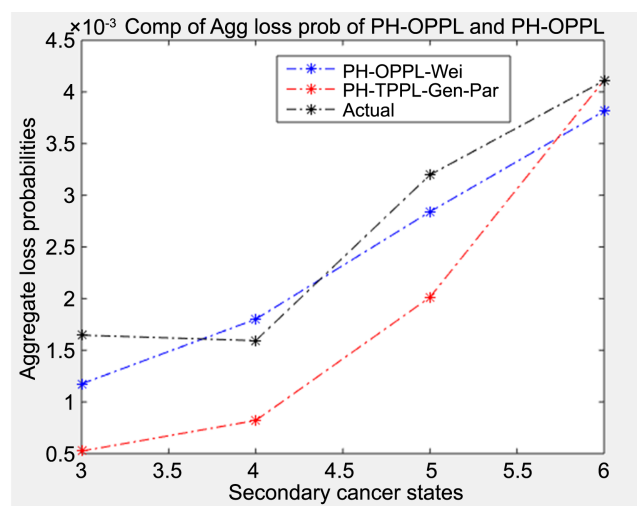
The values of **Table 2** can be represented graphically as:



(a)



(b)



(c)

**Figure 3.** Aggregate loss probabilities.

**Figure 3(a)** shows aggregate loss probabilities using PH-OPPL distribution with severity distributions and it indicates that PH-OPPL with Weibull and Pareto were similar to the actual aggregate loss probabilities while PH-OPPL with generalized Pareto distribution overestimate the aggregate losses for six state model. **Figure 3(b)** shows aggregate loss probabilities using PH-TPPL distribution with Pareto and generalized Pareto provided a better fit for secondary cancer data while PH-TPPL with Weibull overestimated the aggregate losses. However, PH-OPPL with Weibull and PH-TPPL with generalized Pareto provided a better fit compared to PH-OPPL-Pareto model and PH-TPPL Pareto respectively hence they are compared in **Figure 3(c)** indicating that PH-OPPL with Weibull provided the best fit for aggregate loss data of secondary cancers in Kenya. PH-OPPL-Weibull model can be used to provide better estimates of aggregate losses for secondary cancer data in Kenya.

## 5. Conclusion

Mixed phase type distributions are developed to model secondary cancer cases in Kenya. Unlike ordinary distributions which do not in-cooperate the transition of different states, the distributions proposed here take into consideration transition states while modeling claim frequency data. The distributions are based on Poisson and Lindley distributions, where PH-OPPL-Weibull provided the best for PH-OPPL models while PH-TPPL-Generalized Pareto provided the best fit for PH-TPPL models. This model improves estimation of aggregate losses as it in-cooperates transition probabilities of different states of cancer as well as heterogeneous aspect of claim data. This greatly improves estimation of insurance policies for diseases which transit to different state such as cancer hence improving the financial positions of the insurance firms as it will improve estimation of its reserves. This model, however, is only applicable in risk theory for diseases which have multiple transitions states. Further research can be done on this study factoring in patients who were censored in this study and also the same study can be carried out for disease such as HIV-AID which has transition states.

## Data Availability

The data used to support the findings of this study can be availed upon request.

## Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## References

- [1] Shevchenko, P.V. (2010) Calculation of Aggregate Loss Distribution. *Journal of Operational Risk*, **5**, 3-40.
- [2] Robertson, J. (1992) The Computation of Aggregate Loss Distribution.

- [3] Rono, A., Ogutu, C. and Weke, P. (2020) On Compound Distributions for Natural Disaster Modelling in Kenya. *International Journal of Mathematics and Mathematical Sciences*, **2020**, Article ID: 9398309. <https://doi.org/10.1155/2020/9398309>
- [4] Mohamed, M.A., Razali, A.M. and Ismail, N. (2010) Approximation of Aggregate Losses Using Simulation. *Journal of Mathematics and Statistics*, **6**, 233-239. <https://doi.org/10.3844/jmssp.2010.233.239>
- [5] Erlang, A. (1909) Sandsynlighedsregning og telefonsamtaler. *Nyttidsskriftfor Matematik*, **20**, 33-39.
- [6] Neuts, M.F. (1981) Matrix-Geometric Solutions in Stochastic Models. Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore.
- [7] Assussen, S. (2003) Applied Probability and Queues. Springer-Verag, New York.
- [8] Mogens, B. (2005) A Review of Phase-Type Distribution and Their Used in Risk Theory. *Austin Bulletin*, **35**, 145-161. <https://doi.org/10.2143/AST.35.1.583170>
- [9] O’Cinneide, C. (2017) Phase-Type Distributions and Invariant Polytopes.
- [10] Wu, X.Y. and Li, S.M. (2010) Matrix-Form Recursions for a Family of Compound Distributions. *Austin Bulletin*, **40**, 351-368. <https://doi.org/10.2143/AST.40.1.2049233>
- [11] Kok, S. and Wu, X. (2010) Matrix-Form Recursive Evaluation of the Aggregate Claims Distribution Revisited. Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Melbourne.
- [12] Nurul, N.Z., Mahmud, O., Rajalingam, S., Hanita, D., Lazim, A. and Evizal, A.K. (1981) Markov Chain Model Development for Forecasting Air Pollution Index of Miri, Sarawak. *Journal of Sustainability*, **11**, 5190. <https://doi.org/10.3390/su11195190>
- [13] Zhou, Y.J., Wang, L.L., Zhong, R. and Tan, Y.L. (2018) A Markov Chain Based Demand Prediction Model for Stations in Bike Sharing Systems. *Journal of Mathematical Problems in Engineering*, **2018**, Article ID: 8028714. <https://doi.org/10.1155/2018/8028714>
- [14] Das, K.K., Ahmed, I. and Bhattacharjee, S. (2018) A New Three-Parameter Poisson-Lindley Distribution for Modelling Over-Dispersed Count Data. *International Journal of Applied Engineering Research*, **13**, 16468-16477.
- [15] Sankaran, M. (1970) The Discrete Poisson-Lindley Distribution. *Biometrics*, **26**, 145-149. <https://doi.org/10.2307/2529053>
- [16] Shanker, R. and Mishra, A. (2014) A Two-Parameter Poisson-Lindley Distribution. *International Journal of Statistics and Systems*, **9**, 79-85.
- [17] Kemeny, J.G. and Snell, J.L. (2016) Finite Markov Chains. Springer-Verlag, Princeton.