Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 7
July 2023

A Stochastic Model for Three Parameter Generalized Pareto Distribution Using Three Sources

R. Kavitha[1], P. Pandiyan[2], N. Vijayasankar[3]
1. Research scholar, Department of Statistics, Annamalai University, Annamalai Nagar, Tamil Nadu, India.
2. Professor, Department of Statistics, Annamalai University, Annamalai Nagar, Tamil Nadu, India.
3. Assistant professor, Department of Statistics, Govt Arts College, Chidambaram
Corresponding author: kavathir@gmail.com

Abstract

Oral, lung, and stomach cancers are three prevalent and significant malignancies with distinct characteristics, aetiologies, and consequences. Oral cancer primarily affects the oral cavity, including the lips, tongue, gums, and lining of the mouth, and is associated with risk factors such as tobacco and alcohol use, as well as human papillomavirus (HPV) infection. Lung cancer originates in the lung tissues and is mainly linked to tobacco smoking, although exposure to environmental pollutants and genetic factors also contribute to its development. Stomach cancer, also known as gastric cancer, arises from the cells lining the stomach and is influenced by factors such as Helicobacter pylori infection, diet, and genetic predisposition. Metastasis, the spread of cancer cells, is a common occurrence in advanced stages, leading to further complications and poorer prognoses. Genetic mutations, alterations in gene expression, and dysregulation of signalling pathways play essential roles in the initiation and progression of these malignancies. TP53 mutations are frequently observed in all three cancers, while EGFR mutations are prevalent in lung cancer, and CDH1 mutations are found in stomach cancer. Efforts towards early detection, prevention, and personalized treatment strategies are of utmost importance to improve outcomes for patients with oral, lung, and stomach cancers.

Introduction

Cancer is a large group of diseases that can emerge in any organ or tissue of the body during the abnormal or uncontrolled growth of cells. The invasion of the cancer cells to the adjacent parts and its spreading to other body organs is termed metastasis condition, a major reason for mortality. Globally, cancer is the second leading cause of death estimated at around 10 million deaths in 2020. By 2040, the count may rise to 16.4 million.

Lung, prostate, colorectal, stomach, and liver cancer are the most common types of cancer in men, whereas breast, colorectal, lung, cervical, and thyroid cancer are the most common among women (WHO). Worldwide, 60% of cancer-related cases and death are because of eight types of cancers. They are lung, stomach, breast, colon-rectum, mouth, liver, cervix, and oesophagus cancers. Stomach, oral, and lung cancer are three of the most common types of cancer that have significant health consequences and can be fatal if not detected and treated early.

Oral cancer is the 15th most common cancer in men and the 20th most common cancer in women and can affect any part of the mouth. It affects people mostly over 50 years of age. Tobacco use, excessive alcohol use, and human papillomavirus (HPV) infection are the most common risk factors for oral cancer. Tobacco use is the leading cause of oral cancer, accounting for about 75% of cases. The most common genetic changes that are associated with oral cancer are mutations in genes that control cell growth and division. These mutations can lead to uncontrolled cell growth and the formation of tumours. The genes that are commonly mutated in oral cancer are EGFR (epidermal growth factor receptor), p53 (tumour suppressor gene), and KRAS (Kirsten rat sarcoma viral oncogene). The molecular changes that are associated with oral cancer can lead to several changes in the cells, including increased cell growth and division, loss of cell differentiation, increased cell motility, decreased cell apoptosis, etc. Lung cancer is a type of cancer that originates in the lungs. It is the most common type

of cancer occurring worldwide and is one of the leading causes of cancer-related deaths. Smoking is the primary cause of lung cancer and passive smoking, environmental pollutants, or genetic factors are the secondary causes. The potential consequences associated with lung cancer are physical symptoms and functional impairment, metastasis and disease progression, respiratory complications, emotional and psychological impact, treatment-related side effects, and reduced life expectancy. The molecular changes that are associated with Lung cancer are TP53 mutations, EGFR mutations, KRAS mutations, ALK rearrangements, and PD-L1 expression.

Stomach cancer also known as gastric cancer affects the cell lining of the stomach. The various consequences associated with stomach cancer are a range of digestive symptoms including persistent abdominal pain, indigestion, heartburn, nausea, vomiting, bloating, and changes in appetite. Stomach cancer can interfere with the absorption of nutrients from food, leading to malnutrition and deficiencies in essential vitamins and minerals. Advanced stomach cancer can erode blood vessels, leading to gastrointestinal bleeding. Other consequences are perforation and/or blockage of the stomach or intestines. Also, emotional and psychological impact on patients and their families. The molecular changes associated with stomach cancer are CDH1 gene mutations, TP53 mutations, HER2 amplification, Epstein-Barr Virus (EBV) infection, and DNA methylation alterations.

5.2ASSUMPTIONS OF THE MODEL

This assumption is somewhat artificial, but it is made on the one hand due to the lack of detailed information from the real world and on the other hand to describe the production.

- Cigarette Smoking, Tobacco, and alcohol consumption habits are the source of growth for cancer.
- The threshold for each individual is a random variable. If the total damage exceeds the threshold value Y, which in itself is a random variable, habit occurs and a person is recognized as infected.
- If the cumulative damage due to successive events crosses the antigenic diversity threshold level seroconversion takes place. The inter-arrival times between habits, Cigarette Smoking, Tobacco, and alcohol are statistically independent.

5.3 NOTATIONS

X_i : A continuous random variable denoting the amount of contribution to the antigenic diversity due to the alcohol, Tobacco and smoking use in the i^{th} habits, in other words the damage caused to the tumor growth in the i^{th} habits, with p.d.f $g(.)$ and c.d.f $G(.)$.

Y_1, Y_2, Y_3 : continuous random variable denoting the threshold levels for the three components which follows three parameter generalized Pareto distribution.

U_i : a random variable denoting the inter-arrival times between habits with c.d.f. $F_i(.), i = 1,2,3 \dots k$.

$g(.)$: The probability density function of X_i.

$g^*(.)$: Laplace transform of $g(.)$.

$g_k(.)$: The k- fold convolution of $g(.)$ i.e., p.d.f. of $\sum_{j=1}^{k} X_i$

$F_k(.)$: k-fold convolution of $F(.)$.

$f(.)$: p.d.f. of random variable denoting between successive event with the corresponding c.d.f. $F(.)$.

$S(.)$: Survival function. $V_k(t)$: Probability of exactly k habits.

$L(t) : 1 - S(t)$.

5.4 MODEL DESCRIPTION

MODEL DESCRIPTION

$$F(x) = \left[1 - e^{-\left(\frac{x-d}{b}\right)}\right]\left[1 - e^{-\left(\frac{x-d}{b}\right)}\right]\left[1 - e^{-\left(\frac{x-d}{b}\right)}\right]$$

$$= 1 - e^{-\left(\frac{x-d}{b}\right)} - 2e^{-\left(\frac{x-d}{b}\right)} + 2\left[e^{-\left(\frac{x-d}{b}\right)}\right]^2 + 2\left[e^{-\left(\frac{x-d}{b}\right)}\right]^2 - \left[e^{-\left(\frac{x-d}{b}\right)}\right]^3$$

$$= 1 - 3e^{-\left(\frac{x-d}{b}\right)} + 3e^{-2\left(\frac{x-d}{b}\right)} - e^{-3\left(\frac{x-d}{b}\right)}$$

$$\bar{H}(x) = 1 - \left[1 - 3e^{-\left(\frac{x-d}{b}\right)} + 3e^{-2\left(\frac{x-d}{b}\right)} - e^{-3\left(\frac{x-d}{b}\right)}\right]$$

$$= 1 - 1 + 3e^{-\left(\frac{x-d}{b}\right)} - 3e^{-2\left(\frac{x-d}{b}\right)} + e^{-3\left(\frac{x-d}{b}\right)}$$

$$= 3e^{-\left(\frac{x-d}{b}\right)} - 3e^{-2\left(\frac{x-d}{b}\right)} + e^{-3\left(\frac{x-d}{b}\right)}$$

$$\bar{H}(x) = 3e^{\left(\frac{d-x}{b}\right)} - 3e^{2\left(\frac{d-x}{b}\right)} + e^{3\left(\frac{d-x}{b}\right)}$$

$$P(X_i < Y) = \int_0^\infty g_k^*(x)\bar{H}(x)dx$$

$$P(X_i < Y) = \int_0^\infty g^*(x)\bar{H}(x)dx$$

$$P\big((X_i < Y)\big) = \int_0^\infty g^*(x)3e^{\left(\frac{d-x}{b}\right)} - 3e^{2\left(\frac{d-x}{b}\right)} + e^{3\left(\frac{d-x}{b}\right)}dx$$

By taking laplace transform, we get

$$= \int_0^\infty g_k(x)3e^{\left(\frac{d-x}{b}\right)} - 3e^{2\left(\frac{d-x}{b}\right)} + e^{3\left(\frac{d-x}{b}\right)}dx$$

By convolution theorem

Now the threshold Y is such that it has three components namely Y_1, Y_2 and Y_3 component and caner from Y_1 to Y_3 is also possible.

We have the threshold level of seroconversion is given by $Y = max(Y_1, Y_2, Y_3)$.

$$P[max(Y_1, Y_2, Y_3)] = P[(Y_1 < y) \cap (Y_2 < y) \cap (Y_3 < y)] = P[Y_1 < y]P[Y_2 < y]P[Y_3 < y]$$

Now that, Y_1 and Y_2 Y_3follow three parameter generalized Pareto distribution with parameterb, d *and* α .

$$P\left(\sum_{i=1}^{k} X_i < Y\right) = 3\int_0^\infty g^*(x)e^{-\left(\frac{x-d}{b}\right)} - 3\int_0^\infty g^*(x)e^{\left(\frac{2x-2d}{b}\right)} + \int_0^\infty g^*(x)e^{\left(\frac{3x-3d}{b}\right)}dx$$

$$= 3\left[g^*\left(\frac{1-d}{b}\right)\right]^k - 3\left[g^*\left[2\left(\frac{1-d}{b}\right)\right]\right]^k + \left[g^*\left[3\left(\frac{1-d}{b}\right)\right]\right]^k \quad \dots (3)$$

... (1)

... (2)

Probability that the total damage does not cross the threshold level till time t.

S(t) = P(T > t) = Probability that the total damage survives beyond t

$$= \sum_{k=0}^{\infty} \text{P}\{\text{there are exactly k contacts in } (0, t] * \text{P (the total cumulative } (0, t]\}$$

$$S(t) = P(T > t) = \sum_{k=0}^{\infty} V_k(t)P\big(X_i < max(Y_1, Y_2, Y_3)\big) \dots (4)$$

It may happen that successive shocks become increasingly effective in causing damage, even though they are independent. This means that $V_k(t)$ is the probability and that there are exactly k occasions of damage in (o,t).

P(exactly k contacts in $(0, t]$) = $F_k(t) - F_{K+1}(t)$ with $F_0(t) = 1$

$$= \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\left(\frac{1-d}{b}\right)\right]^k + \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\left(\frac{1-d}{b}\right)\right]^k - \sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\,2\left(\frac{1-d}{b}\right) + \left(\frac{1-d}{b}\right)\right]^k \ldots (5)`$$

$L(t) = 1 - S(t)$, Taking laplace transform of $L(t)$, We get

$$L(t) = 1 - \left\{3\sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\left(\frac{1-d}{b}\right)\right]^k - 3\sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\left(\frac{2-2d}{b}\right)\right]^k + 3\sum_{k=0}^{\infty}[F_k(t) - F_{k+1}(t)]\left[g^*\left(\frac{3-3d}{b}\right)\right]^k\right\} \ldots \quad (6)$$

On simplification we get,

$$L(t) = 3\left[1 - g^*\left(\frac{1-d}{b}\right)\right]\sum_{k=1}^{\infty}F_k(t)\left[g^*\left(\frac{1-d}{b}\right)\right]^{k-1} - 3\left[1 - g^*\left(\frac{2-2d}{b}\right)\right]\sum_{k=1}^{\infty}F_k(t)\left[g^*\left(\frac{2-2d}{b}\right)\right]^{k-1} + \left[1 - g^*\left(\frac{3-3d}{b}\right)\right]\sum_{k=1}^{\infty}F_k(t)\left[g^*\left(\frac{3-3d}{b}\right)\right]^{k-1} \ldots$$

Bytaking Laplace-Stieltjes transform, it can be shown that

$$l^*(s) = \frac{3\left[1 - g^*\left(\frac{1-d}{b}\right)\right]f^*(s)}{\left[1 - g^*\left(\frac{1-d}{b}\right)f^*(s)\right]} - \frac{3\left[1 - g^*\left(\frac{2-2d}{b}\right)\right]f^*(s)}{\left[1 - g^*\left(\frac{2-2d}{b}\right)f^*(s)\right]} + \frac{\left[1 - g^*\left(\frac{3-3d}{b}\right)\right]f^*(s)}{\left[1 - g^*\left(\frac{3-3d}{b}\right)f^*(s)\right]} \ldots$$

Let the random variable U denoting inter arrival time which follows exponential with parameter$c$. Now $f^*(s) = \left(\frac{c}{c+s}\right)$, substituting in the above equation (8) we get

$$= \frac{3\left[1 - g^*\left(\frac{1-d}{b}\right)\right]\left(\frac{c}{c+s}\right)}{\left[1 - g^*\left(\frac{1-d}{b}\right)\left(\frac{c}{c+s}\right)\right]} - \frac{3\left[1 - g^*\left(\frac{2-2d}{b}\right)\right]\left(\frac{c}{c+s}\right)}{\left[1 - g^*\left(\frac{2-2d}{b}\right)\left(\frac{c}{c+s}\right)\right]} + \frac{\left[1 - g^*\left(\frac{3-3d}{b}\right)\right]\left(\frac{c}{c+s}\right)}{\left[1 - g^*\left(\frac{3-3d}{b}\right)\left(\frac{c}{c+s}\right)\right]} \ldots \quad (9)$$

$$E(T) = -\frac{d}{ds}l^*(s) given\ s = 0, \qquad E(T^2) = \frac{d^2}{ds^2}l^*(s)\ given\ s = 0 \quad (7)$$

From which variance $V(T) = E(T^2) - [E(T)]^2$can be obtained

$$\mathrm{E}(T) = \frac{3}{c\left[1 - g^*\left(\frac{1-d}{b}\right)\right]} - \frac{3}{c\left[1 - g^*\left(\frac{2-2d}{b}\right)\right]} + \frac{1}{c\left[1 - g^*\left(\frac{3-3d}{b}\right)\right]}$$ on simpilification

$$E(T) = \frac{3b}{c[(1-d+\mu b)]} - \frac{3b}{c[2(1-d)+\mu b]} + \frac{b}{c[3(1-d)+\mu b]}$$ on simpilification (10)

$$E(T^2) = \frac{3}{c^2\left[1-g^*\left(\frac{1-d}{b}\right)\right]^2} - \frac{3}{c^2\left[1-g^*\left(\frac{2-2d}{b}\right)\right]^2} + \frac{1}{c^2\left[1-g^*\left(\frac{3-3d}{b}\right)\right]^2}$$

$$E(T^2) = \frac{3b^2}{c^2[(1-d+\mu b)]^2} - \frac{3b^2}{c^2[2(1-d)+\mu b]^2} + \frac{b^2}{c^2[3(1-d)+\mu b]^2}$$

$$V(T) = E(T^2) - (E(T))^2$$

$$= \left[\frac{3b^2}{c^2[(1-d+\mu b)]^2} - \frac{3b^2}{c^2[2(1-d)+\mu b]^2} + \frac{b^2}{c^2[3(1-d)+\mu b]^2}\right] - \left[\frac{3b}{c[(1-d+\mu b)]} - \frac{3b}{c[2(1-d)+\mu b]} + \frac{b}{c[3(1-d)+\mu b]}\right]^2$$

$$V(T) = \frac{18b^2}{C^2[1-d+\mu b][2(1-d)+\mu b]} + \frac{6b^2}{C^2[2(1-d)+\mu b][3(1-d)+\mu b]} - \frac{6b^2}{C^2[1-d+\mu b][3(1-d)+\mu b]} - \frac{6b^2}{C^2[1-d+\mu b]^2} - \frac{6b^2}{C^2[2(1-d)+\mu b]^2}$$ .... (11)

Table : 1 infected person's stage wise

C	b	μ	d	E(t)	V(t)
1	13	9	8	0.111321	0.095636
2	14	10	6	0.050017	0.017398
3	18	11	4	0.030304	0.005859
4	16	16	5	0.015625	0.001561
5	18	16	6	0.0125	0.001006
6	17	15	5	0.011111	0.00079
7	19	14	8	0.010205	0.000696
8	22	18	6	0.006945	0.000305
9	21	18	8	0.006173	0.000247
10	26	14	5	0.007143	0.00032

Fig : 1: The Chart for infected person's Expected time

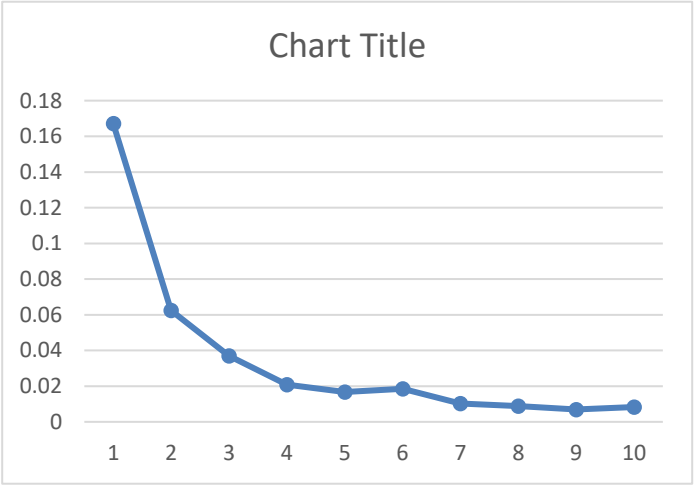
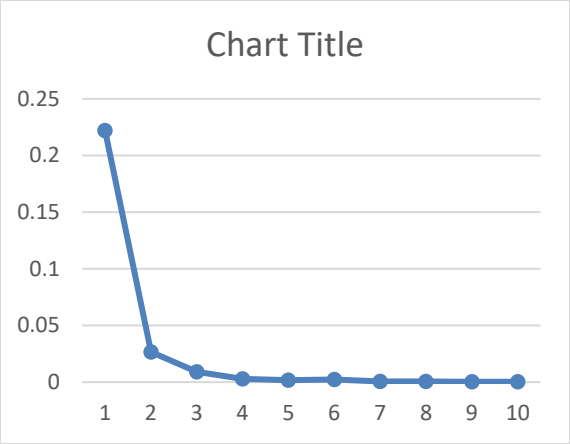


Fig 2: The Chart for infected person's variance time



CONCLUSION

In conclusion, a person with oral cancer who has FEV1, FVC cell infection crosses the threshold faster. Both men and women are likely to be impacted if infected with tumour cells because once the individual is infected, the FEV1 and FVC cell development are destroyed and the projected lifetime rapidly declines until it reaches the threshold value. The length of the infection depends on the duration that the sick individual was exposed to smoke. The table and pictures demonstrate how the immune system fails after infection, as demonstrated by the model. If a person is diagnosed with cancer, the proliferation of healthy cells is observed and FEV1, FVC cells are damaged by the mouth cancer as a result decrease in immunity was observed.

As a result, in all scenarios of parameter values, the anticipated time E (T) value to exceed the threshold of cancer patients is lowering. As seen in the Figure, the predicted time decreases if the TP53, CDH1, EGFR parameter's value rises. The similar impact may be seen in the variance V (T) finding for cancer patients as shown in the figure. According to the graph, the lifespan of drinkers, smokers, and stomach cancer patients is getting shorter with time. Therefore, getting the right guidance from doctors and receiving regular care might lengthen his or her life.

Rreferences

1. Vivek Borse,∗ Aditya Narayan Konwar, and Pronamika Buragohain(2020). Oral cancer diagnosis and perspectives in India, Sensors International 1 (2020) 100046.
2. Krishna A, Singh S, Kumar V, Pal US.(2015). Molecular concept in human oral cancer. Natl J Maxillofac Surg. Vol.6(1):PP 9-15.
3. Polanski J, Jankowska-Polanska B, Rosinczuk J, Chabowski M, Szymanska-Chabowska A(2016). Quality of life of patients with lung cancer. Onco Targets Ther, 29;9:1023-8.
4. Hamzehloie T, Mojarrad M, Hasanzadeh Nazarabadi M, Shekouhi S. (2012).The role of tumor protein 53 mutations in common human cancers and targeting the murine double minute 2-p53 interaction for cancer therapy. Iran J Med Sci.37(1):3-8. PMID: 23115424; PMCID: PMC3470295.
5. Goulding RE, Chenoweth M, Carter GC, Boye ME, Sheffield KM, John WJ, Leusch (2020). mutation as a prognostic factor and predictive factor in advanced/metastatic non-small cell lung cancer: A systematic literature review and meta-analysis. Cancer Treat Res Commun.24:100200. doi: 10.1016/j.ctarc.2020.100200. Epub 2020 Jul 25. PMID: 32750661.
6. Chang, GC., Yang, TY., Chen, KC(2020). *ALK* variants, PD-L1 expression, and their association with outcomes in *ALK*-positive NSCLC patients. *Sci Rep* 10, 21063.
7. Cencioni C, Trestini I, Piro G, Bria E, Tortora G, Carbone C, Spallotta F (2022). Gastrointestinal Cancer Patient Nutritional Management: From Specific Needs to Novel Epigenetic Dietary Approaches. Nutrients. 8;14(8):1542. doi: 10.3390/nu14081542. PMID: 35458104; PMCID: PMC9024975.
8. Liyen Cartelle A, Uy PP, Yap JEL. Acute Gastric Hemorrhage due to Gastric Cancer Eroding Into a Splenic Artery Pseudoaneurysm: Two Dangerously Rare Etiologies of Upper Gastrointestinal Bleeding. Cureus. 2020 Sep 27;12(9):e10685. doi:

10.7759/cureus.10685. PMID: 33133851; PMCID: PMC7593120.

9. Sahli H, Mandour JE, Tessi RTY, Jerguigue H, Latib R, Omor Y. An unusual cause of peritonitis: Perforation of a gastric carcinoma. Radiol Case Rep. 2021 Dec 28;17(3):740-743. doi: 10.1016/j.radcr.2021.11.055. PMID: 35003472; PMCID: PMC8717435.
10. Benton B, Norton C, Lindsay J, et al. Can nurses manage gastrointestinal symptoms arising from pelvic radiation disease? *Clin Oncol* 2011; 23:538–51
11. Paez, J.G. et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science, 304(5676), 1497-1500.
12. *Patrick Tan, Khay-Guan Yeoh(2015). Genetics and Molecular Pathogenesis of Gastric AdenocarcinomaGastroenterology, Vol 149, Issue 5, PP 1153-1162.e3.*
13. Pandiyan P and Koventhan S. Estimating of survival time of cancer Patients, International Journal of Recent Scientific Research, 9(1), 23533 – 23534, (2018).
14. Pandiyan P. Calculating Human Immune Virus Infected Person Survival Time Through Stochastic Model, Bio- Science Research Bulletin, 29(1): 1 - 4(2013).
15. Ammar Sarhan and Zaindin M. Modified Weibull distribution, Applied Sciences, 11: 123 – 136, (2009).
16. P. Pandiyan, S. Koventhan, G. Subash Chandra bose Estimating The Survival Time Of cancer Drinker Patients Using Stochastic Model, Indo – Asian Journal of Multidisciplinary Research Volume – 4(4); 1464 – 1467 ,(2018)
17. S. Jothimanickam & P. Pandiyan (2019), Calculating The Survival Time Of Cancer Patients Through Exponentiated Weibull Distribution, Journal of International and Computational Science,Volume 9, Issue 12.
18. S. Jothimanickam, P. Pandiyan, E. Susiganeshkumar & T. Vignesh (2020), Stochastic Model for Expected Time Using Generalized Exponential Distribution,Journal Of Xi'an University Of Architecture &Technology, Volume XII, Issue IX, September 2020.
19. Saksena.S.K and Johnson A.M, (1984). "Best unbiased estimators for the parameters Pareto distribution". Bio-metrika, Vol.31, pp.77-83.