E-ISSN: 3025-0994
Halaman 1251-1264
Volume 4 Nomor 4 Tahun 2026

Seleksi Model Generalized Linear untuk Data Count dengan Overdispersion dan Excess Zeros pada Kasus Penyakit di Kabupaten Gresik Tahun 2020–2025

Albert Servant Ndruru[1], Alfon Brian Harefa [2], Femi Ezrani Sianturi[3], Finny Aprilia Yulanda[4], Widi Aulia[5]
[1,2,3,4,5]Program Studi Statistika, Universitas Negeri Medan
e-mail: albertservantndruru@gmail.com, alfonharefa2005@gmail.com, femiezrani@gmail.com, finyyulanda@gmail.com, auliawidi237@gmail.com

Abstrak

Data jumlah kasus penyakit merupakan data cacah yang dapat memiliki varians lebih besar daripada rata-rata dan proporsi nilai nol yang tinggi. Kondisi tersebut dapat menyebabkan model Poisson kurang sesuai dalam menggambarkan karakteristik data sehingga diperlukan pemilihan model yang tepat. Penelitian ini bertujuan mengidentifikasi overdispersi dan kelebihan nilai nol serta menentukan model Generalized Linear yang paling sesuai untuk memodelkan jumlah kasus HIV/AIDS, Demam Berdarah Dengue, diare, tuberkulosis, dan malaria di Kabupaten Gresik tahun 2020–2025. Data yang digunakan merupakan data sekunder dengan 108 observasi yang berasal dari 18 kecamatan selama enam tahun. Analisis dilakukan dengan membandingkan model Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Hurdle-Poisson, dan Hurdle-Negative Binomial. Pemilihan model didasarkan pada Akaike Information Criterion, Bayesian Information Criterion, Likelihood Ratio Test, dan Vuong Test. Hasil penelitian menunjukkan bahwa kasus HIV/AIDS, Demam Berdarah Dengue, diare, dan tuberkulosis mengalami overdispersi, sedangkan malaria tidak menunjukkan overdispersi yang signifikan. Model Negative Binomial menghasilkan nilai Akaike Information Criterion dan Bayesian Information Criterion terendah untuk seluruh jenis penyakit. Hasil pengujian juga menunjukkan bahwa model zero-inflated tidak lebih unggul dibandingkan model standar. Dengan demikian, model Negative Binomial merupakan model yang paling sesuai untuk memodelkan jumlah kasus penyakit dalam penelitian ini.

Kata kunci: Data Cacah; Overdispersi; Nilai Nol; Negative Binomial; Kasus Penyakit

Abstract

Disease case counts are count data that may have variances greater than their means and high proportions of zero observations. These characteristics can make the Poisson model less suitable for representing the data and require appropriate model selection. This study aimed to identify overdispersion and excess zeros and to determine the most

appropriate Generalized Linear model for modeling HIV/AIDS, dengue hemorrhagic fever, diarrhea, tuberculosis, and malaria cases in Gresik Regency from 2020 to 2025. Secondary data consisting of 108 observations from 18 subdistricts over six years were used. The analysis compared the Poisson, Negative Binomial, Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Hurdle-Poisson, and Hurdle-Negative Binomial models. Model selection was based on the Akaike Information Criterion, Bayesian Information Criterion, Likelihood Ratio Test, and Vuong Test. The results showed that HIV/AIDS, dengue hemorrhagic fever, diarrhea, and tuberculosis cases exhibit overdispersion, while malaria does not show significant overdispersion. The Negative Binomial model has the lowest Akaike Information Criterion and Bayesian Information Criterion values for all diseases. The results also indicate that zero-inflated models do not outperform the corresponding standard models. Therefore, the Negative Binomial model is the most appropriate model for describing disease case counts in this study.

Keywords : count data; overdispersion; excess zeros; Negative Binomial; disease cases

LATAR BELAKANG

Penyakit menular masih menjadi salah satu permasalahan kesehatan yang perlu mendapatkan perhatian karena jumlah kasus dapat berbeda antarwilayah dan mengalami perubahan dari waktu ke waktu. Informasi mengenai jumlah kasus penyakit pada tingkat wilayah penting untuk menggambarkan pola kejadian penyakit serta mendukung perencanaan dan evaluasi program kesehatan. Dalam analisis statistik, jumlah kasus penyakit umumnya berbentuk data cacah (*count data*), yaitu data yang memiliki nilai berupa bilangan bulat tidak negatif. Data penelitian ini mencakup jumlah kasus HIV/AIDS, Demam Berdarah Dengue (DBD), diare, tuberkulosis (TB), dan malaria pada 18 kecamatan di Kabupaten Gresik selama tahun 2020–2025 dengan total 108 observasi.

Regresi Poisson merupakan salah satu metode dalam *Generalized Linear Model* (GLM) yang dapat digunakan untuk memodelkan data cacah. Dalam regresi Poisson terdapat asumsi *equidispersion*, yaitu nilai rata-rata dan varians data harus memiliki nilai yang sama. Namun, dalam penerapannya asumsi tersebut sering tidak terpenuhi karena varians data dapat lebih besar daripada nilai rata-ratanya. Kondisi tersebut disebut *overdispersion*. Ambarwati et al. (2020) menjelaskan bahwa regresi Poisson memiliki asumsi *equidispersion*, sedangkan ketika varians lebih besar daripada rata-rata maka terjadi *overdispersion*. Kondisi tersebut dapat menyebabkan model Poisson kurang sesuai apabila digunakan tanpa penanganan terhadap dispersi data.

Permasalahan *overdispersion* dapat ditangani dengan menggunakan regresi Negative Binomial. Regresi Negative Binomial memiliki parameter dispersi tambahan yang memungkinkan varians model lebih besar daripada nilai rata-ratanya. Ulfa et al. (2021) membandingkan regresi Poisson dan Negative Binomial pada jumlah kasus baru kusta di Pulau Jawa dan menunjukkan bahwa pendekatan Negative Binomial dapat digunakan untuk menangani *overdispersion* pada regresi Poisson. Penelitian Rafifah et

al. (2022) pada kasus DBD di Jawa Barat juga menunjukkan bahwa data mengalami *overdispersion* sehingga regresi Negative Binomial memberikan hasil yang lebih baik dibandingkan regresi Poisson berdasarkan nilai AIC dan BIC.

Penerapan regresi Negative Binomial juga banyak digunakan pada pemodelan kasus penyakit di Indonesia. Ariani dan Widodo (2018) menggunakan regresi Negative Binomial untuk menganalisis faktor-faktor yang memengaruhi jumlah penderita DBD di Jawa Tengah. Hasil penelitian menunjukkan bahwa regresi Negative Binomial merupakan model yang lebih baik dibandingkan regresi Poisson ketika data mengalami *overdispersion*. Banapon et al. (2020) juga menerapkan regresi Negative Binomial pada kasus tuberkulosis di Jawa Barat dan menjelaskan bahwa model tersebut digunakan untuk menangani pelanggaran asumsi *equidispersion* pada regresi Poisson. Selain itu, Aipassa et al. (2023) menggunakan regresi Negative Binomial untuk menganalisis jumlah penyakit kusta di Provinsi Maluku karena data penyakit tersebut merupakan data cacah yang berpotensi mengalami *overdispersion*.

Meskipun regresi Negative Binomial dapat menangani *overdispersion*, karakteristik data cacah tidak selalu hanya ditandai oleh varians yang lebih besar daripada rata-rata. Data kasus penyakit juga dapat memiliki proporsi nilai nol yang tinggi atau *excess zeros*. Kondisi tersebut dapat menjadi pertimbangan dalam pemilihan model karena nilai nol dapat muncul akibat proses pembentukan data yang berbeda dari observasi dengan nilai positif. Oleh karena itu, selain model Poisson dan Negative Binomial, model Zero-Inflated Poisson (ZIP), Zero-Inflated Negative Binomial (ZINB), serta model Hurdle dapat dipertimbangkan untuk data yang memiliki banyak nilai nol. Pada penelitian mengenai malaria di Sumatera Utara, Dalimunthe dan Husein (2025) menggunakan Zero-Inflated Negative Binomial (ZINB) untuk memodelkan kasus malaria yang memiliki karakteristik nilai nol dan dispersi yang perlu diperhatikan.

Karakteristik *overdispersion* dan *excess zeros* menunjukkan bahwa pemilihan model untuk data cacah perlu dilakukan secara sistematis. Penggunaan satu model secara langsung tanpa memeriksa karakteristik data dapat menghasilkan model yang kurang sesuai. Oleh karena itu, beberapa model dapat dibandingkan berdasarkan ukuran kecocokan dan kompleksitas model seperti Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC). Selain itu, pengujian statistik dapat digunakan untuk menentukan apakah parameter dispersi maupun komponen nilai nol diperlukan dalam model. Dalam penelitian ini, enam model dibandingkan, yaitu Poisson, Negative Binomial (NB), Zero-Inflated Poisson (ZIP), Zero-Inflated Negative Binomial (ZINB), Hurdle-Poisson (HP), dan Hurdle-Negative Binomial (HNB).

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menyeleksi model Generalized Linear yang sesuai dalam memodelkan jumlah kasus HIV/AIDS, DBD, diare, TB, dan malaria menurut kecamatan di Kabupaten Gresik selama periode 2020–2025. Penelitian ini bertujuan untuk mengidentifikasi keberadaan *overdispersion* dan *excess zeros*, membandingkan beberapa model regresi data cacah, serta menentukan model terbaik berdasarkan AIC, BIC, Likelihood Ratio Test (LRT), dan Vuong test. Dengan adanya seleksi model tersebut, diharapkan diperoleh model yang

mampu menggambarkan karakteristik jumlah kasus penyakit di Kabupaten Gresik secara lebih tepat.

METODE

Penelitian ini menggunakan data sekunder mengenai jumlah kasus HIV/AIDS, Demam Berdarah Dengue (DBD), diare, tuberkulosis (TB), dan malaria di Kabupaten Gresik selama periode 2020–2025. Data kasus diperoleh dari Gresik Satu Data/Dinas Kesehatan Kabupaten Gresik dan mencakup 18 kecamatan dengan total 108 observasi. Data jumlah penduduk menurut kecamatan dan tahun diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Gresik dan digunakan sebagai *offset* dalam pemodelan.

Sebelum dilakukan analisis, data melalui tahap preparasi untuk memastikan konsistensi dan kelengkapan data. Preparasi meliputi perbaikan penulisan nama wilayah, yaitu mengubah penulisan “C e r m e” menjadi “Cerme”, serta pemeriksaan terhadap nilai yang hilang. Berdasarkan hasil pemeriksaan data, terdapat 4 nilai kosong pada HIV/AIDS, 4 nilai kosong pada DBD, dan 31 nilai kosong pada malaria. Seluruh nilai kosong pada malaria berada pada periode 2020–2021. Nilai yang tidak dilaporkan tidak langsung dianggap sebagai nol karena data yang tidak tersedia tidak dapat dipastikan menunjukkan tidak adanya kasus. Oleh karena itu, analisis utama mempertahankan nilai tersebut sebagai *missing value*. Selanjutnya dilakukan analisis sensitivitas dengan skenario alternatif, yaitu memperlakukan nilai yang tidak dilaporkan sebagai nol, untuk mengetahui apakah perubahan perlakuan terhadap data tersebut memengaruhi hasil pemodelan.

Analisis dilakukan secara terpisah untuk setiap jenis penyakit menggunakan enam model regresi data cacah, yaitu regresi Poisson, *Negative Binomial* (NB), *Zero-Inflated Poisson* (ZIP), *Zero-Inflated Negative Binomial* (ZINB), *Hurdle-Poisson*, dan *Hurdle-Negative Binomial* (HNB). Variabel respons berupa jumlah kasus penyakit, sedangkan prediktor yang digunakan adalah tahun dan kecamatan. Jumlah penduduk digunakan sebagai *offset* untuk memperhitungkan perbedaan ukuran populasi antarwilayah. Parameter model diestimasi menggunakan *Maximum Likelihood Estimation* (MLE).

Model Poisson digunakan sebagai model dasar untuk data cacah dengan asumsi bahwa nilai harapan dan varians memiliki nilai yang sama. Model Poisson dirumuskan sebagai

$$Y_{it} \sim Poisson(\mu_{it})$$

dengan fungsi peluang

$$P(Y_{it} = y_{it}) = \frac{e^{-\mu_{it}} \mu_{it}^{y_{it}}}{y_{it}!}, \qquad y_{it} = 0,1,2,\cdots$$

dengan Y_{it} merupakan jumlah kasus pada kecamatan ke-i dan tahun ke-t, sedangkan μ_{it} merupakan nilai harapan jumlah kasus.

Model dengan *offset* jumlah penduduk dinyatakan dalam bentuk:

$$\log(\mu_{it}) = \beta_0 + \beta_1 Tahun_{it} + \gamma_i \log(Populasi_{it})$$

dengan β_0 sebagai intersep, β_1 sebagai koefisien tahun, γ_i sebagai efek kecamatan, dan $Populasi_{it}$ sebagai jumlah penduduk pada kecamatan i dan tahun t.

Model NB digunakan apabila terdapat *overdispersion*, yaitu kondisi ketika varians jumlah kasus lebih besar daripada nilai rata-ratanya. Bentuk varians model NB dinyatakan sebagai

$$Var\,(Y_{it}) = \mu_{it} + \alpha\mu_{it}^2, \quad \alpha > 0$$

dengan α merupakan parameter dispersi.

Model ZIP dan ZINB digunakan untuk mengakomodasi kondisi *excess zeros*, yaitu proporsi nilai nol yang lebih besar daripada yang dapat dijelaskan oleh model Poisson atau NB biasa. Model ZIP menggabungkan proses pembentukan nilai nol dengan distribusi Poisson, sedangkan ZINB menggabungkan proses *zero inflation* dengan distribusi Negative Binomial. Secara umum, untuk model ZIP,

$$P(Y_{it} = 0)\pi_{it} + (1 - \pi_{it})e^{-\mu_{it}}$$

sedangkan untuk $y_{it} > 0$,

$$P(Y_{it} = y_{it}) = (1 - \pi_{it})\frac{e^{-\mu_{it}}\mu_{it}^{y_{it}}}{y_{it}!}$$

dengan π_{it} merupakan probabilitas suatu observasi berasal dari proses *zero inflation*.

Model *Hurdle-Poisson* dan *Hurdle-Negative Binomial* digunakan sebagai alternatif untuk memisahkan proses terbentuknya nilai nol dari proses banyaknya kasus positif. Pada model Hurdle, bagian pertama memodelkan probabilitas terjadinya kasus atau tidak terjadinya kasus, sedangkan bagian kedua memodelkan jumlah kasus positif menggunakan distribusi yang terpotong pada nol (*zero-truncated distribution*). *Hurdle-Poisson* menggunakan distribusi Poisson pada bagian positif, sedangkan *Hurdle-Negative Binomial* menggunakan distribusi Negative Binomial.

Parameter seluruh model diestimasi menggunakan *Maximum Likelihood Estimation* (MLE). Secara umum, estimator maksimum likelihood diperoleh dengan memaksimalkan fungsi

$$\hat{\theta} = \arg\max_{\theta} L(\theta; y)$$

Pemilihan model dilakukan dengan membandingkan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC). AIC dirumuskan sebagai

$$AIC = -2\ell\left(\hat{\theta}\right) + 2k$$

sedangkan BIC dirumuskan sebagai

$$BIC = -2\ell\left(\hat{\theta}\right) + 2k\log(n)$$

dengan $\ell\left(\hat{\theta}\right)$ merupakan nilai *log-likelihood* maksimum, k jumlah parameter model, dan n jumlah observasi. Model dengan nilai AIC dan BIC yang lebih rendah dipertimbangkan memiliki keseimbangan yang lebih baik antara kecocokan dan kompleksitas model.

Perbandingan model Poisson dan NB dilakukan menggunakan *Likelihood Ratio Test* (LRT) untuk mengevaluasi kebutuhan parameter dispersi tambahan. Statistika LRT dinyatakan sebagai:

$$LR = \left[2\ell\left(\hat{\theta}_{NB}\right) - \ell\left(\hat{\theta}_{P}\right)\right]$$

Perbandingan model Poisson dengan ZIP serta NB dengan ZINB dilakukan menggunakan *Vuong test*. Kecocokan model juga dievaluasi secara visual menggunakan *rootogram* dengan membandingkan distribusi frekuensi data aktual dan frekuensi yang diharapkan berdasarkan model. Model akhir dipilih berdasarkan kombinasi AIC, BIC, hasil pengujian model, dan kecocokan distribusi data. Analisis dilakukan menggunakan perangkat lunak R dengan bantuan paket MASS, pscl, glmmTMB, dan DHARMa.

HASIL DAN PEMBAHASAN

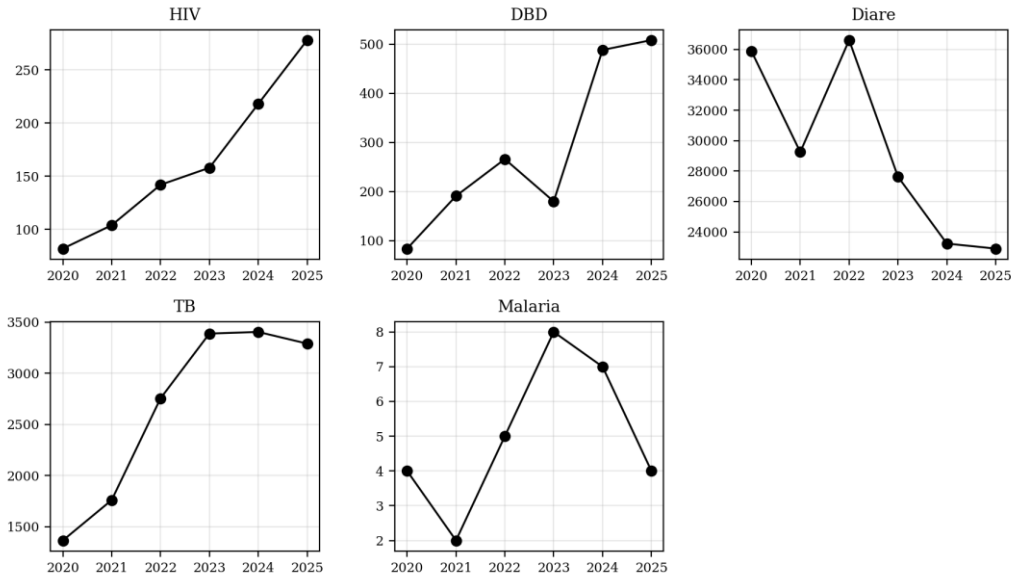
Data yang dianalisis terdiri atas 108 observasi panel (18 kecamatan × 6 tahun, 2020–2025) untuk lima penyakit, yaitu HIV/AIDS, demam berdarah dengue (DBD), diare, tuberkulosis (TB), dan malaria. Sel kosong pada data HIV/AIDS dan DBD tahun 2020 serta malaria tahun 2020–2021 diperlakukan sebagai nol kasus. Ringkasan statistik deskriptif disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif Jumlah Kasus per Kecamatan-Tahun

Penyakit	Min–Maks	Rata-rata	Varians	Var/Mean	Nol (%)
HIV/AIDS	0–69	9,09	91,56	10,07	9 (8,33)
DBD	0–88	15,90	307,68	19,35	9 (8,33)
Diare	120–4.375	1.625,11	1.046.975,00	644,25	0 (0,00)
TB	15–899	147,80	18.288,89	123,74	0 (0,00)
Malaria	0–5	0,28	0,46	1,67	86 (79,63)

Sumber: data diolah. Nol (%) = jumlah dan persentase observasi bernilai nol.

Tabel 1 menunjukkan bahwa varians jauh melampaui rata-rata pada empat penyakit pertama (rasio varians terhadap rata-rata 10,07 untuk HIV/AIDS hingga 644,25 untuk diare), yang mengindikasikan overdispersi. Malaria memiliki rata-rata terendah (0,28 kasus) dengan 79,63% observasi bernilai nol, sedangkan diare dan TB tidak memiliki observasi bernilai nol. Gambar 1 memperlihatkan total kasus per tahun: HIV/AIDS meningkat dari 82 kasus (2020) menjadi 278 kasus (2025), DBD dari 84 menjadi 508 kasus, dan TB dari 1.365 menjadi 3.290 kasus, sedangkan diare menurun dari 35.890 menjadi 22.900 kasus.



Gambar 1. Total kasus per tahun menurut penyakit, Kabupaten Gresik 2020–2025

Overdispersi diuji dengan uji Cameron–Trivedi pada model regresi Poisson dengan prediktor tahun dan kecamatan (Tabel 2).

Tabel 2. Hasil Uji Overdispersi pada Model Poisson

Penyakit	Dispersi Pearson	z	p-value	Keputusan
HIV/AIDS	1,994	2,853	0,002	Overdispersi
DBD	4,447	4,844	<0,001	Overdispersi
Diare	191,180	6,483	<0,001	Overdispersi
TB	18,345	3,335	<0,001	Overdispersi
Malaria*	1,663	1,131	0,129	Tidak signifikan

*Malaria: prediktor hanya tahun, karena banyak kecamatan bernilai nol pada seluruh periode sehingga koefisien kecamatan tidak stabil (quasi-separation).

Pada taraf signifikansi 5%, overdispersi terbukti signifikan untuk HIV/AIDS, DBD, diare, dan TB, sedangkan untuk malaria tidak signifikan (p = 0,129). Dari sisi nol berlebih, proporsi nol HIV/AIDS dan DBD (masing-masing 8,33%) memang jauh di atas proporsi yang diharapkan distribusi Poisson (≈0,01% dan ≈0,00%). Namun, selisih tersebut belum cukup untuk menyimpulkan adanya inflasi nol karena overdispersi sendiri dapat menghasilkan lebih banyak nol. Proporsi nol malaria (79,63%) hanya sedikit di atas harapan Poisson (75,75%).

Enam model dibandingkan, yaitu Poisson, binomial negatif (NB), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), hurdle Poisson (HP), dan hurdle negative binomial (HNB). Model terbaik ditentukan berdasarkan nilai AIC dan BIC terkecil (Tabel 3).

Tabel 3. Nilai AIC dan BIC Model Count

Penyakit	Kriteria	Poisson	NB	ZIP	ZINB	HP	HNB
HIV/AIDS	AIC	602,12	**585,89**	600,92	588,11	610,79	598,19
	BIC	653,09	**639,53**	657,24	647,11	667,12	657,20
DBD	AIC	846,12	**688,96**	844,68	691,97	874,04	720,72
	BIC	897,08	**742,60**	901,00	750,98	930,36	779,73
Diare	AIC	18.122,30	**1.653,24**	–	–	–	–
	BIC	18.173,26	**1.706,88**	–	–	–	–
TB	AIC	2.380,49	**1.086,99**	–	–	–	–
	BIC	2.431,45	**1.140,63**	–	–	–	–
Malaria*	AIC	155,05	**149,03**	153,51	152,91	153,53	152,50
	BIC	160,41	**157,08**	164,24	166,32	164,26	165,91

Angka tebal = nilai terkecil. Tanda "–" = model tidak dapat diestimasi karena data tidak memuat nilai nol. *Malaria: prediktor hanya tahun, sehingga nilainya tidak dapat dibandingkan langsung dengan penyakit lain.

Model NB memiliki AIC dan BIC terkecil pada kelima penyakit. Pada diare dan TB, model dengan komponen nol (ZIP, ZINB, HP, HNB) tidak dapat diestimasi karena tidak ada observasi bernilai nol, sehingga komponen inflasi nol tidak teridentifikasi. Pada HIV/AIDS, ZINB (AIC 588,11) mendekati NB (585,89), tetapi BIC tetap memilih NB. Untuk malaria, uji rasio likelihood Poisson terhadap NB menghasilkan $\chi^2 = 8{,}02$ ($p < 0{,}01$), sejalan dengan AIC dan BIC yang memilih NB. Hal ini menunjukkan bukti overdispersi malaria lebih lemah dibanding penyakit lain karena uji Cameron–Trivedi tidak signifikan.

Keberadaan inflasi nol diuji lebih lanjut dengan uji Vuong pada Tabel 4.

Tabel 4. Hasil Uji Vuong Model Zero-Inflated terhadap Model Standar

Penyakit	Pembanding	z (Raw)	p (Raw)	z (BIC)	p (BIC)
HIV/AIDS	ZIP vs Poisson	0,905	0,183	−0,724	0,235
	ZINB vs NB	0,567	0,285	−2,418	0,008
DBD	ZIP vs Poisson	0,835	0,202	−0,603	0,273
	ZINB vs NB	0,454	0,325	−3,859	<0,001
Malaria*	ZIP vs Poisson	0,794	0,214	−0,550	0,291
	ZINB vs NB	0,185	0,427	−14,139	<0,001

z positif memihak model zero-inflated; z negatif memihak model standar (Poisson/NB). Raw = tanpa koreksi; BIC = koreksi BIC. *Malaria: prediktor hanya tahun.

Statistik Vuong tanpa koreksi tidak signifikan pada seluruh pasangan model ($p > 0,05$), sehingga model zero-inflated tidak terbukti lebih baik daripada model standarnya. Dengan koreksi BIC, nilai z negatif pada semua pasangan dan signifikan untuk ZINB terhadap NB pada HIV/AIDS ($p = 0,008$), DBD ($p < 0,001$), dan malaria ($p < 0,001$), yang berarti NB lebih disukai. Dengan demikian, keragaman berlebih pada data kasus di Kabupaten Gresik lebih tepat ditangani dengan binomial negatif daripada dengan komponen inflasi nol.

Estimasi parameter model NB disajikan pada Tabel 5 untuk efek tahun dan Tabel 6 untuk efek kecamatan, dengan Kecamatan Balongpanggang sebagai kategori referensi. Nilai incidence rate ratio (IRR) diperoleh dari eksponen koefisien.

Tabel 5. Estimasi Efek Tahun dan Parameter Dispersi Model NB

Penyakit	Model	β Tahun	SE	p-value	IRR Tahun	θ (SE)
HIV/AIDS	NB	0,2417	0,0268	<0,001	1,273	13,60 (5,23)
DBD	NB	0,3475	0,0322	<0,001	1,415	5,13 (1,04)
Diare	NB	−0,1261	0,0182	<0,001	0,882	9,64 (1,30)
TB	NB	0,1867	0,0160	<0,001	1,205	14,35 (2,18)
Malaria*	NB	0,1100	0,1319	0,404	1,116	0,566 (0,340)

θ = parameter dispersi NB (semakin kecil, semakin besar overdispersi). *Malaria: prediktor hanya tahun.

Tabel 6. IRR Kecamatan terhadap Balongpanggang pada Model NB

Kecamatan	HIV/AIDS	DBD	Diare	TB
Balongpanggang (ref.)	1,000	1,000	1,000	1,000
Benjeng	0,798	0,819	0,937	0,777
Bungah	0,701	0,724	0,634*	0,986
Cerme	1,794*	0,885	1,199	1,100
Driyorejo	1,618	0,529*	1,957***	1,329
Duduksampeyan	0,769	0,480*	0,765	0,554***
Dukun	0,965	0,609	1,046	0,629**
Gresik	2,737***	1,840*	1,727**	3,105***
Kebomas	3,909***	3,132***	1,511*	1,816***
Kedamean	1,054	0,508*	0,587**	0,589**
Manyar	1,617	2,612***	2,204***	1,661**
Menganti	1,577	1,099	2,023***	1,349
Panceng	0,618	0,584	0,348***	0,398***
Sangkapura	0,210***	0,029***	1,229	0,273***
Sidayu	0,754	0,684	0,522***	0,376***
Tambak	0,064***	0,028***	0,178***	0,189***
Ujungpangkah	0,687	0,875	0,692*	0,662*
Wringinanom	0,405**	0,632	1,108	0,530***

*p < 0,05; **p < 0,01; ***p < 0,001. Malaria tidak ditampilkan karena modelnya tidak memuat kecamatan.

Efek tahun signifikan pada empat penyakit pertama (p < 0,001). Dengan kecamatan dianggap tetap, kasus HIV/AIDS meningkat rata-rata 27,3% per tahun (IRR = 1,273), DBD 41,5% (IRR = 1,415), dan TB 20,5% (IRR = 1,205), sedangkan diare menurun 11,8% per tahun (IRR = 0,882). Pada malaria, efek tahun tidak signifikan (IRR = 1,116; p = 0,404), sehingga tidak ada bukti perubahan tren.

Antarkecamatan, Kebomas dan Gresik konsisten memiliki IRR tinggi. Kebomas menunjukkan kasus HIV/AIDS 3,909 kali dan DBD 3,132 kali Balongpanggang, sedangkan Gresik memiliki kasus TB 3,105 kali dan HIV/AIDS 2,737 kali. Manyar tinggi untuk DBD (2,612), diare (2,204), dan TB (1,661). Sebaliknya, Tambak dan Sangkapura memiliki IRR sangat rendah pada HIV/AIDS, DBD, dan TB (misalnya IRR DBD 0,028 dan 0,029), dengan pengecualian diare di Sangkapura yang tidak berbeda nyata dari referensi.

Hasil seleksi menunjukkan bahwa persoalan utama pada data ini adalah overdispersi, bukan inflasi nol. Overdispersi yang besar pada diare (rasio varians terhadap rata-rata 644,25) dan TB (123,74) membuat model Poisson sangat tidak memadai, yang tercermin dari AIC Poisson untuk diare (18.122,30) yang jauh di atas NB (1.653,24). Temuan ini sejalan dengan Hayati dkk. (2026) yang mendapati overdispersi pada data kasus TB di Provinsi Lampung sehingga regresi binomial negatif dipilih berdasarkan AIC. Nabilah dan Sofro (2026) juga melaporkan bahwa pada kasus diare tingkat kabupaten/kota di Jawa Barat, rasio dispersi model Poisson mencapai 1.819 dan binomial negatif menjadi model terbaik menurut AIC dan BIC, tanpa indikasi inflasi nol pada diagnostik residualnya. Untuk HIV/AIDS dan DBD, Maulana dan Ariyanto (2026) juga menggunakan regresi binomial negatif untuk mengatasi overdispersi pada kasus ODHIV di 38 kabupaten/kota Jawa Timur, dan Fadil dkk. (2024) melaporkan bahwa model yang sama mengatasi overdispersi pada kasus DBD di Kota Makassar. Kesamaan pola pada empat penyakit tersebut memperkuat bahwa binomial negatif adalah titik awal yang wajar untuk data kasus penyakit tingkat wilayah.

Untuk malaria, hasil penelitian ini berbeda dengan Dalimunthe dan Husein (2025) yang menemukan bahwa ZINB lebih sesuai daripada Poisson dan NB pada kasus malaria di Sumatera Utara tahun 2022. Masruroh dan Ariyanto (2026) juga memilih model zero-inflated untuk kasus difteri di Jawa Timur tahun 2024 yang memiliki 76,32% nol dan dispersi 1,44, tetapi penelitian itu hanya membandingkan ZIP dan ZINB sehingga tidak menguji apakah komponen inflasi nol memang diperlukan dibanding model standar. Perbedaan dengan hasil Gresik mungkin dipengaruhi oleh karakteristik data, seperti proporsi nol dan cakupan wilayah, sehingga perlu dikaji lebih lanjut. Pada data Gresik, proporsi nol malaria (79,63%) hanya sedikit di atas harapan Poisson (75,75%) dan uji Vuong tidak mendukung inflasi nol. Hal ini konsisten dengan simulasi Kurnia dkk. (2026) yang menunjukkan bahwa model zero-inflated unggul terutama ketika jumlah nol sangat tinggi, sedangkan binomial negatif menghasilkan estimasi dispersi yang mendekati nilai ideal.

Pola spasial memperlihatkan konsentrasi kasus di kecamatan Kebomas, Gresik, dan Manyar, sedangkan Tambak dan Sangkapura, yang berada di Pulau Bawean, konsisten rendah. Karena model ini tidak menyertakan jumlah penduduk sebagai offset, perbedaan IRR antarkecamatan sebagian dapat mencerminkan perbedaan ukuran populasi, tidak sepenuhnya perbedaan risiko. Nabilah dan Sofro (2026) menggunakan offset jumlah penduduk agar koefisien menggambarkan laju insidensi, dan pendekatan serupa dapat diterapkan pada penelitian lanjutan di Gresik. Jumlah kasus yang tercatat juga dapat dipengaruhi oleh ketersediaan dan akses layanan pemeriksaan di masing-masing wilayah.

Terdapat beberapa keterbatasan. Pertama, tahun dimodelkan secara linear pada skala log, sedangkan Gambar 1 menunjukkan pola yang tidak sepenuhnya linear (DBD turun pada 2023 dan TB mendatar sejak 2023), sehingga IRR tahun merupakan rata-rata laju perubahan. Kedua, sel kosong 2020 yang dianggap nol dapat menaikkan tren HIV/AIDS dan DBD secara artifisial. Ketiga, model malaria hanya memuat tahun sehingga tidak dapat menangkap perbedaan antarkecamatan.

SIMPULAN

Hasil penelitian menunjukkan bahwa data kasus HIV/AIDS, DBD, diare, dan tuberkulosis mengalami overdispersi, sedangkan inflasi nol tidak terbukti memerlukan model zero-inflated. Berdasarkan nilai AIC dan BIC, model Negative Binomial merupakan model terbaik untuk kelima jenis penyakit. Uji Vuong juga menunjukkan bahwa model zero-inflated tidak lebih unggul dibandingkan model standar. Hasil pemodelan menunjukkan peningkatan kasus HIV/AIDS, DBD, dan tuberkulosis, penurunan kasus diare, serta tidak adanya perubahan tahunan yang signifikan pada malaria. Dinas Kesehatan Kabupaten Gresik disarankan menggunakan model Negative Binomial sebagai dasar analisis dan perencanaan pengendalian penyakit, dengan mempertimbangkan perbedaan karakteristik antarkecamatan.

DAFTAR PUSTAKA

Aipassa, A. D., Wattimena, A. Z., & Haumahu, G. 2023. Analisis Faktor-Faktor yang Mempengaruhi Jumlah Penyakit Kusta di Provinsi Maluku dengan Menggunakan Regresi Binomial Negatif. *Parameter: Jurnal Matematika, Statistika dan Terapannya*, 2(2): 87–100.

Alwansyah, M. A., Uswah, A., & Fauzi, Y. 2025. Comparison of geographically weighted generalized Poisson regression (GWGPR) and geographically weighted negative binomial regression (GWNBR) methods in determining factors affecting tuberculosis cases in Indonesia. *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, 18(1): 851–865. https://doi.org/10.36456/jstat.vol18.no1.a10073

Ambarwati, P. C., Indahwati, & Aidi, M. N. 2020. Kajian Simulasi Overdispersi pada Regresi Poisson dan Binomial Negatif Terboboti Geografis untuk Data Balita Gizi Buruk. *Indonesian Journal of Statistics and Its Applications*, 4(3): 484–497. https://doi.org/10.29244/ijsa.v4i3.684

Ananda, E. Y. P., Annas, S., Ihsan, H., Sukarna, & Aswi. 2024. Negative Binomial Regression Analysis of Factors Influencing Stunting Cases in Central Lombok Regency. *Inferensi*, 7(3): 167–175. https://doi.org/10.12962/j27213862.v7i3.21436

Ariani, P. M., & Widodo, E. 2018. Analisis Faktor Penyebab Penyakit DBD di Jawa Tengah Menggunakan Regresi Binomial Negatif. *Jurnal Kesehatan Vokasional*, 3(1): 1–6. https://doi.org/10.22146/jkesvo.33870

Banapon, A., Laugiwa Prawira Putra, M., & Widodo, E. 2020. Penerapan Regresi Binomial Negatif untuk Mengatasi Pelanggaran Overdispersi pada Regresi Poisson (Studi Kasus Penderita Tuberculosis di Provinsi Jawa Barat Tahun 2017). *BIAStatistics Journal of Statistics Theory and Application*, 14(1): 53–63. https://doi.org/10.1234/bias.v14i1.95

Campbell, H. 2021. The consequences of checking for zero-inflation and overdispersion in the analysis of count data. *Methods in Ecology and Evolution*, 12(4): 665–680.

Dalimunthe, R. A., & Husein, I. 2025. Zero inflated negative binomial regression in malaria cases in North Sumatera. *JISTech (Journal of Islamic Science and Technology)*, 10(1): 107–115. https://doi.org/10.30829/jistech.v10i1.25658

Deni, R., & Desmita, Z. 2016. Perbandingan Model Regresi Generalized Poisson dan Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson (Studi Kasus: Penderita Filariasis di Provinsi Riau Tahun 2011). *Jurnal Sains Matematika dan Statistika*, 2(2): 1–11. https://doi.org/10.24014/jsms.v2i2.3133

Fadil, M., Raupong, R., & Ilyas, N. 2024. Mengatasi overdispersi menggunakan regresi binomial negatif dengan penaksir maksimum likelihood pada kasus demam berdarah di Kota Makassar. *ESTIMASI: Journal of Statistics and Its Application*, 5(1): 37–51. https://doi.org/10.20956/ejsa.v5i1.14552

Feng, C. X. 2021. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. *Journal of Statistical Distributions and Applications*, 8: Article 8. https://doi.org/10.1186/s40488-021-00121-4

Green, J. A. 2021. Too many zeros and/or highly skewed? A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression. *Health Psychology and Behavioral Medicine*, 9(1): 436–455. https://doi.org/10.1080/21642850.2021.1920416

Hayati, M., Sari, R. P., Azizah, R., Hudori, M., & Sari, K. 2026. An analysis of overdispersed tuberculosis case data in Lampung Province using negative binomial. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 14(1): 54–65. https://doi.org/10.24252/msa.v14i1.63646

Ibnas, R., Satriani, S., & Nurfadilah, K. 2023. Negative binomial and generalized Poisson regression model for death due to dengue hemorrhagic fever data. *Eigen Mathematics Journal*, 6(1): 39–48. https://doi.org/10.29303/emj.v6i1.153

Kurnia, A., Fakhriyah, Z., Sadik, K., & Handayani, D. 2026. On the use of zero-inflated mixed models for count data: A simulation and empirical evidence. *Journal of Applied Data Sciences*, 7(2): 1545–1557. https://doi.org/10.47738/jads.v7i2.1303

Lee, K. H., Pedroza, C., Avritscher, E. B. C., Mosquera, R. A., & Tyson, J. E. 2023. Evaluation of negative binomial and zero-inflated negative binomial models for the analysis of zero-inflated count data: Application to the telemedicine for children with medical complexity trial. *Trials*, 24: Article 613. https://doi.org/10.1186/s13063-023-07648-8

Masruroh, I. D., & Ariyanto, D. 2026. Perbandingan model regresi zero-inflated Poisson (ZIP) dan regresi zero-inflated negative binomial (ZINB) terhadap kasus difteri di Provinsi Jawa Timur tahun 2024. *MATHunesa: Jurnal Ilmiah Matematika*, 14(2): 423–431. https://doi.org/10.26740/mathunesa.v14n02.p423-431

Maulana, I., & Ariyanto, D. 2026. Pemodelan jumlah kasus HIV di Provinsi Jawa Timur menggunakan regresi binomial negatif. *MATHunesa: Jurnal Ilmiah Matematika*, 14(2): 490–499. https://doi.org/10.26740/mathunesa.v14n02.p490-499

Muntafiah, R., Wasono, R., & Darsyah, M. Y. 2014. Pemodelan Regresi Binomial Negatif untuk Mengatasi Overdispersion pada Regresi Poisson. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 2(1): 40–45.

Mustafidah, M. E., & Sofro, A. 2024. Pemodelan jumlah kematian akibat filariasis di Jawa Timur dengan menggunakan regresi Hurdle Negative Binomial. *MATHunesa: Jurnal Ilmiah Matematika*, 12(2): 428–438. https://doi.org/10.26740/mathunesa.v12n2.p428-438

Nabilah, N., & Sofro, A. 2026. Negative binomial modeling of district-level diarrhea incidence in West Java Province. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, 11(2): 1451–1471. https://doi.org/10.18860/cauchy.v11i2.42057

Nketia, K., & de Souza, D. K. 2024. Using zero-inflated and hurdle regression models to analyze schistosomiasis data of school children in the southern areas of Ghana. *PLOS ONE*, 19(7): Article e0304681. https://doi.org/10.1371/journal.pone.0304681

Rafifah, Z. 2022. *Penerapan Regresi Binomial Negatif untuk Mengatasi Overdispersi pada Regresi Poisson Kasus Demam Berdarah di Jawa Barat*. Skripsi. Bogor: IPB University. https://repository.ipb.ac.id/handle/123456789/113602

Sandjadirja, L. M., Aidi, M. N., & Rizki, A. 2019. Penanganan Overdispersi pada Regresi Poisson dengan Regresi Binomial Negatif pada Kasus Kemiskinan di Indonesia. *Xplore: Journal of Statistics*, 8(1). https://doi.org/10.29244/xplore.v8i1.165

Shinami, M. A., Hafiyusholeh, M., Wardani, S. A., Khaulasari, H., & Fanani, A. 2026. Model Regresi Binomial Negatif dalam Menganalisis Faktor-Faktor yang Mempengaruhi Jumlah Penderita Diare. *MATHunesa: Jurnal Ilmiah Matematika*, 14(1): 70–80. https://doi.org/10.26740/mathunesa.v14n1.p70-80

Tiara, Y., Aidi, M. N., Erfiani, & Rachmawati, R. 2023. Overdispersion handling in Poisson regression model by applying negative binomial regression. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 17(1): 417–426.

Tinambunan, T. J. B., Nufus, N. H., Meilawati, N. L., Rahma, R., & Wicaksono, F. 2025. Application of the Geographically Weighted Negative Binomial Regression (GWNBR) Method to Tuberculosis Cases in North Sumatra Province in 2024. *Proceedings of the International Conference on Data Science and Official Statistics*, 2025(1): 1119–1133. https://doi.org/10.34123/icdsos.v2025i1.474

Ulfa, Y. A., Soleh, A. M., & Sartono, B. 2021. Handling of Overdispersion in the Poisson Regression Model with Negative Binomial for the Number of New Cases of Leprosy in Java: Penanganan Overdispersi pada Model Regresi Poisson dengan Binomial Negatif untuk Jumlah Kasus Baru Kusta di Jawa. *Indonesian Journal of Statistics and Its Applications*, 5(1): 1–13. https://doi.org/10.29244/ijsa.v5i1p1-13

Yıldırım, G., Kaçıranlar, S., & Yıldırım, H. 2022. Poisson and negative binomial regression models for zero-inflated data: An experimental study. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, 71(2): 601–615. https://doi.org/10.31801/cfsuasmas.988880