Hurdle Models in Psychology—A Practical Guide for Inflated Data

Article *in* International Journal of Psychology · April 2025

DOI: 10.1002/ijop.70042

Renaud Mabire
Lumière University Lyon 2

International Journal of Psychology

EMPIRICAL ARTICLE

Hurdle Models in Psychology—A Practical Guide for Inflated Data

Renaud Mabire-Yon

PSITEC Laboratory – Psychology: Interactions, Time, Emotions, Cognition, University of Lille, Lille, France

Correspondence: Renaud Mabire-Yon (renaud.mabire.yon@gmail.com)

Received: 23 August 2023 | **Revised:** 3 March 2025 | **Accepted:** 26 March 2025

Funding: The author received no specific funding for this work.

Keywords: count data analysis | data science in psychology | Hurdle model | psychological research | R programming | zero inflation

ABSTRACT

In psychological research, variables often exhibit point-mass inflation—for example, many zero responses or other boundary lumps—that defy standard regression techniques. Hurdle models address this challenge by separating the zero-generating process from the distribution of nonzero (or non-boundary) observations, thereby allowing for more accurate modelling of behaviour and outcomes. In this paper, I introduce the conceptual basis of Hurdle models and demonstrate how they can be applied to count data as well as other types of data (e.g., continuous variables with excess zeros). Using a step-by-step tutorial in R, I illustrate how the two-part hurdle structure—consisting of a binary component for point-mass observations and a truncated distribution for positive (or above-threshold) values—provides nuanced insights that simpler models often miss. To illustrate this approach, I walk through a fictional dataset examining home-based HIV testing among men who have sex with men, highlighting the Hurdle model's ability to simultaneously handle overdispersion and excess zeros. Emphasising iterative model evaluation, goodness-of-fit checks and a series of practical recommendations, this paper aims to equip psychologists with a robust analytical framework that promotes deeper, theory-aligned interpretations of data—ultimately fostering innovative research in diverse areas of psychological science.

1 | Introduction

1.1 | Why Point-Mass Inflations Matter in Psychology

In psychological research, variables often display an overabundance of observations at a single point. Although this phenomenon frequently manifests as 'zero inflation' (e.g., many participants reporting zero occurrences of a behaviour), it also appears at other boundaries, such as maximum scores on a questionnaire. When conventional statistical approaches—like standard linear or Poisson regression—are applied to such distributions, they fail to capture the processes generating the point mass or the intensity of the behaviour among those exceeding that threshold (Vives et al. 2006).

1.2 | Hurdle Models as a Two-Part Solution

A flexible way to address this issue is the Hurdle model, which separates the process by which observations remain at the inflated point (often zero) from the process that governs variation among those who cross that threshold. Hurdle models have historically been applied to zero-inflated count data (Cragg 1971; Mullahy 1986; Heilbron 1994), but they can just as well handle other distributions.

1.3 | Scope of This Paper

In this article, we use a fictional example to illustrate the basic principles of Hurdle's model: Dr. Efigenia Nasa's study of home-based HIV testing among men who have sex with men (MSM). Thus, we demonstrate how to identify when a Hurdle model is warranted, how to implement it in R, and how to interpret results. *The model is implemented using simulated data, and the code needed to generate it is available as additional material.* Finally, we suggest solutions to the most common pitfalls encountered in implementing this model.

2 | Historical and Theoretical Background

2.1 | The Original Developments

The Hurdle model emerged from a series of statistical innovations designed to address the limitations of traditional count models. The seminal work can be traced back to Cragg (1971), who proposed a 'two-part' model of household expenditure, separating the decision to purchase from the amount spent. This concept laid the foundation for modelling processes with separate decision stages. Building on this foundation, Mullahy (1986) formally introduced the Hurdle model in the context of count data analysis, providing a more flexible way of dealing with excess zeros than standard count models such as Poisson regression or negative binomial (NB) regression.

2.2 | Extensions to Non-Count Data

Heilbron (1994) further developed the theoretical foundations of Hurdle models, exploring their applications in epidemiology and health services research. While historically applied to zero-inflated count data, Hurdle models can also be extended to other types of outcomes (Greene 1994; Brooks et al. 2017), including continuous variables with excess zeros (Hurdle Gamma) or proportions bounded between 0 and 1 (Hurdle Beta). This broader applicability aligns with many psychological studies where an 'inflated' value arises from a distinct data-generating mechanism.

3 | Why and When to Use the Hurdle Model?

3.1 | Zero Inflation Versus Point-Mass Inflation

Zero inflation[1] refers to the presence of more observed zeros than conventional models predict (Lambert 1992). However, point-mass inflation extends beyond zero to include, for example, a high incidence of maximum scores or minimum nonzero values. In psychology, these inflations might stem from distinct processes—for instance, a portion of participants who never adopt a behaviour or who always select the top scale category. Hurdle models address this by explicitly modelling the 'hurdle-crossing' decision separately from variation among those who cross it.

3.2 | Zero-Inflated Versus Hurdle Versus Standard Models

When deciding how to model count data, researchers often begin with standard approaches such as Poisson or NB regression (Hilbe 2011). These methods are usually appropriate when the mean–variance relationship is appropriate (the Poisson requires mean ≈ variance, while the NB accommodates moderate overdispersion), and the zeros in the data set do not exceed what these distributions predict. However, an excess of zeros—one that markedly surpasses the Poisson or NB expectation—indicates a potential mismatch (Lambert 1992). In these cases, two main classes of 'zero-augmented' models can address the problem:

- *Zero-inflated models* (*ZIP*, *zero-inflated negative binomial* [*ZINB*]): These models assume that zeros arise from *two* sources (Lambert 1992): (1) a 'sure-zero' or *structural* process, for individuals incapable of producing a positive count and (2) the usual Poisson or NB count process for everyone else, which can also generate zeros by chance. Then, zero-inflated models can complicate interpretation by requiring the analyst to distinguish 'structural zeros' from 'sampling zeros' (Loeys et al. 2012). *For details on the different types of zeros, see* Box 1.
- *Hurdle models*: By contrast, hurdle (or 'two-part') models posit one source of zeros: failure to cross the 'hurdle'. If a participant does not surpass this binary gate, they remain at zero; if they do, their positive count follows a truncated distribution (Mullahy 1986). This simplifies interpretation when all individuals are *capable* of producing a nonzero value, but some never do so (Heilbron 1994).

To help you find your way through the model selection process, see Figure 1.

BOX 1 | Structural versus sampling zeros.

In count data modelling, it is common to define structural zeros as arising from individuals who *cannot* produce a positive outcome under any circumstance (Lambert 1992; Heilbron 1994). By contrast, sampling zeros are thought to reflect a scenario in which an event *could* have occurred, yet did not—due to chance, behavioural choices, or timing (Cragg 1971; Mullahy 1986). This distinction underpins the development of:

- Zero-inflated models (ZIP, ZINB), which explicitly model a subgroup of 'sure-zero' participants—those genuinely incapable of engagement—plus a regular count process for everyone else.
- Hurdle models, which assume *all* zeros result from a single binary hurdle—everyone in principle can produce a positive count, but some opt (or happen) not to cross that threshold (Loeys et al. 2012).

Why misclassification can bias results

Treating structural zeros as if they were sampling zeros (or vice versa) can lead to biased estimates and misinterpretation of regression coefficients. For example, if a truly 'impossible' subset is lumped together with those who simply choose zero, the model's zero-generation process becomes confounded, inflating or deflating the effects of certain predictors (Lambert 1992). This, in turn, can lead to misalignment with theory: if the study posits a subgroup that is absolutely unable to engage, but the analysis forces everyone into a single 'opt-out' mechanism, the empirical findings may contradict the underlying psychological or behavioural framework.

Psychosocial barriers and ambiguous zero

Although 'cannot' is often interpreted literally (e.g., a physical or legal impossibility), psychosocial factors can blur the line between 'cannot' and 'did not'. An individual who has nominal access to an HIV self-test, for example, may face such significant stigma or personal constraints that testing feels impossible. One could classify this as a *sampling zero* (there is some chance, however minimal) or treat it as *functionally structural* if the likelihood is effectively zero in that social context. The choice depends on:

1. *Theoretical rationale*: Does prior research frame these psychosocial hurdles as absolute barriers or as extreme deterrents still on a continuum of possibility?
2. *Empirical indicators*: Are there participants in similar conditions who nonetheless engage, suggesting the behaviour is *theoretically* feasible, although low probability?

Practical suggestions

- *Match the model to the conceptual reality*: If you have strong evidence of a truly impossible subgroup, zero-inflated models are logical. Otherwise, a hurdle framework may suffice.
- *Incorporate relevant predictors*: Covariates capturing stigma, resource constraints, or psychosocial stressors help distinguish near-zero probabilities from actual impossibility.
- *Stay empirical and iterative*: If in doubt, compare zero-inflated versus hurdle specifications via model fit criteria (AIC, BIC) and interpret them in light of your theoretical assumptions (Loeys et al. 2012).

3.3 | Overdispersion and the Hurdle Approach

A frequent complication in count data is overdispersion, in which the variance exceeds the mean (Hilbe 2011). Although NB models handle moderate overdispersion, an excess of zeros can push variance far beyond the scope of a single-process count model (Coxe et al. 2009).

Hurdle models mitigate overdispersion by separating the zero-generating mechanism from the positive counts. Once the zeros are modelled in a dedicated logistic (or probit) component, the truncated distribution for positive values may exhibit less variance relative to the mean (Wang 2010). In many psychological datasets, this resolves the majority of overdispersion tied to high zero counts (Loeys et al. 2012). If the remaining positive counts *still* display significant overdispersion, analysts can opt for a hurdle NB instead of a Hurdle Poisson (Hilbe 2011).

In short:

- *Hurdle Poisson*: Works well if, after removing zeros in a separate process, the positive counts are approximately Poisson-like.
- *Hurdle NB*: Needed when the positive counts themselves remain overdispersed relative to the mean.

4 | Hurdle Model Functioning

4.1 | Technical Structure (Two-Part Model)

A Hurdle model divides the data-generating process into two components (Mullahy 1986; Heilbron 1994).

- *Binary component*: A logistic or probit regression determines whether an observation is zero versus nonzero. Anyone who 'fails' to cross this hurdle has a zero count.
- *Truncated count component*: A Poisson or NB distribution, truncated to exclude zero, models only the positive counts. By focusing on observations above zero, this component estimates frequency or intensity among those who have already engaged.

Together, these two equations provide a coherent statistical framework. In psychology, this approach aligns with the idea that 'initiation' of a behaviour (the zero vs. nonzero decision) can differ from the amount of that behaviour (the truncated count).

4.2 | Comparison With Zero-Inflated Frameworks

Zero-inflated models (ZIP, ZINB) also use two processes, but they split zeros between structural and sampling sources (Lambert 1992). Specifically, in a ZINB model, zeros arise from both a 'sure-zero' mechanism and the NB distribution, whereas a Hurdle NB funnels all zeros through a logistic hurdle and only allows positive counts to follow the NB distribution. Interpreting which zeros arise from true inability can be cumbersome, especially if everyone in the sample is technically capable of producing a positive count but many choose not to (Loeys et al. 2012). By contrast, Hurdle models treat all zeros as sampling zeros: individuals could engage but do not cross the hurdle for reasons captured in the binary model (Mullahy 1986; Feng 2021). This avoids the structural versus sampling distinction, often simplifying interpretation. If theory suggests that some participants truly cannot produce a positive count, zero-inflated models may be more suitable. Otherwise, the hurdle assumption—'everyone can do it; some simply don't'—is frequently a better conceptual fit.

4.3 | Other Distributions

While Hurdle models are most familiar in zero-inflated count scenarios, the same logic applies whenever a single value is inflated, even for continuous or bounded data: (a) Hurdle Gamma, for continuous outcomes with a mass at zero (e.g., daily minutes of

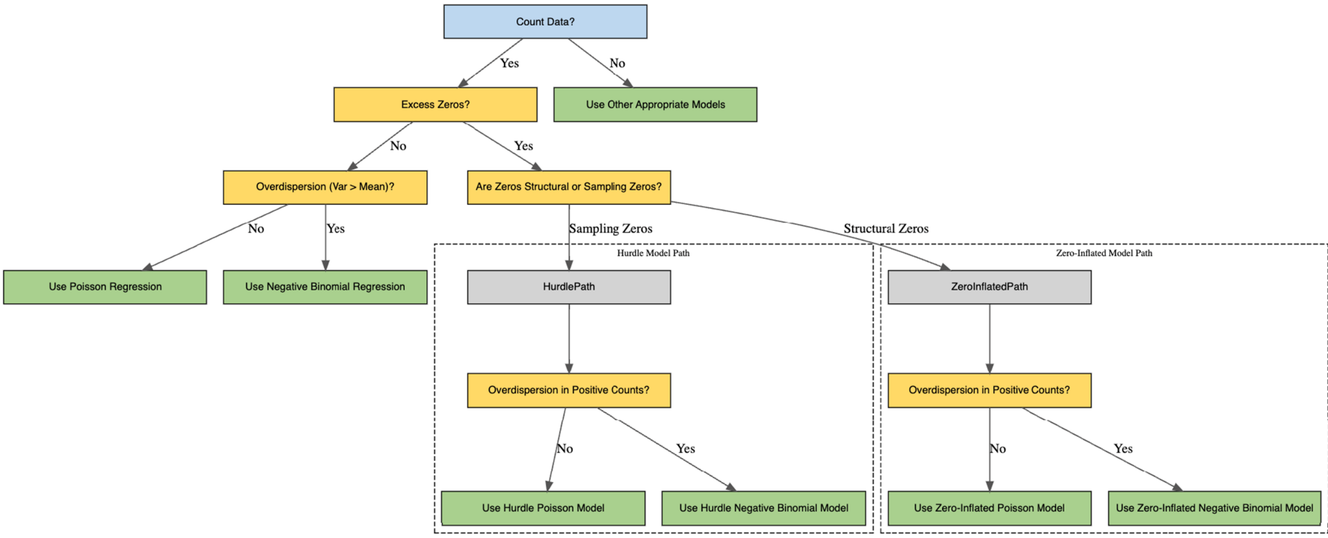


FIGURE 1 | Navigating model selection for count data analysis. This figure (the decision tree) illustrates how researchers might choose between these options, based on whether zeros appear 'structural', whether overdispersion remains an issue, and how the data align with Poisson or NB assumptions (Greene 1994). *Note*: For continuous outcomes with point-mass inflation, a similar conceptual approach applies (e.g., Hurdle Gamma), although this specific diagram targets count data.

exercise), (b) Hurdle Beta, for proportion data, possibly inflated at 0, 1 or both (Brooks et al. 2017) and (c) Hurdle Gaussian, less common, but feasible if zero or another specific value is overrepresented in a variable otherwise assumed to be normally distributed (Heilbron 1994).

In all these cases, the binary part models whether the observation is at the point mass, and a truncated distribution captures variation beyond that point. Packages like glmmTMB (Brooks et al. 2017) facilitate these extended hurdle families, making it possible to apply hurdle logic to a wide range of psychological data.

5 | Illustrative Example: Zero-Inflated Count Data (R Tutorial)

5.1 | Context and Dataset

Dr. Efigenia Nasa, a health psychology researcher, aimed to investigate the factors influencing the uptake and frequency of at-home HIV testing among MSM. She distributed free HIV self-testing kits to a sample of 1300 MSM and followed them for 6 months. Upon examining the frequency distribution of test usage (Figure 2), Dr. Efigenia Nasa observed a substantial number of participants reporting no test use (zero count), while among those who did use tests, the frequency varied considerably.

Because the data suggested both an excess of zeros and potential overdispersion, Dr. Efigenia Nasa identified the Hurdle model as an analytical approach. The data we will use to explain the implementation of the Hurdle model have been simulated. They, therefore, do not reflect any empirical reality. The fictitious study that produced the data concerned the frequency of HIV testing among 1300 young MSM.

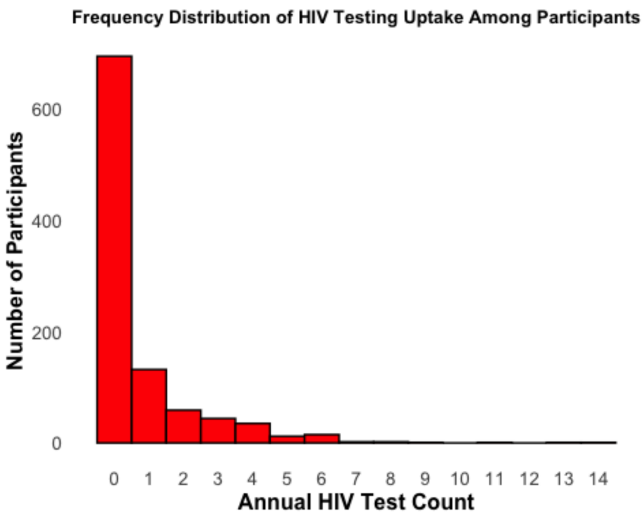


FIGURE 2 | Distribution of HIV testing among participants.

Variables included:

- *Age*: 20–30 years, targeting the young MSM demographic.
- *Education level*: Binary (0 = high school or less, 1 = beyond high school).
- *Perceived risk*: Continuous (0–1), individual's perceived HIV risk.
- *Access to healthcare*: Binary (0 = poor, 1 = good access).
- *Sexual behaviour risk*: Continuous (0–1), level of perceived risk of sexual behaviour.

- *Social support*: Continuous (0 to 1), extent of social support.
- *HIV tests*: Count of HIV tests taken in the past year.

The dataset simulates (see additional materiel for the code generating the data) the decision-making process for HIV testing and the test frequency among those opting for testing. Variables such as perceived risk, access to healthcare, sexual behaviour risk, and social support inform these simulations.

```
# View the first few rows of the dataframe

head(simulated_data)

##   age education_level perceived_risk access_healthcare sexual_behavior_risk
## 1  20               0     0.82540961                 0            0.3382830
## 2  27               0     0.08536057                 0            0.8449936
## 3  21               1     0.63112563                 1            0.8334973
## 4  21               1     0.71007743                 0            0.6955449
## 5  27               0     0.04044810                 0            0.7228493
## 6  25               1     0.60341377                 0            0.2526943
##   social_support tests_HIV
## 1      0.9219214         1
## 2      0.3406063         5
## 3      0.8370911         2
## 4      0.8723637         0
## 5      0.8371167         2
## 6      0.7429962         3
```

5.2 | Preparing Data for Hurdle Modelling

5.2.1 | Initial Steps

Before applying the Hurdle model, it is vital to prepare the dataset meticulously. This preparation involves two primary steps: outlier detection and a thorough data check to ensure the dataset's compatibility with Hurdle model assumptions.

5.2.2 | Outlier Detection and Handling

The detect_outliers_loop function (see additional material for the function) plays a role in identifying and managing outliers. Outliers can skew the analysis, leading to inaccurate model predictions. This function systematically identifies outliers using the interquartile range (IQR) method and provides options to remove, keep, or mark these data points, based on a user-defined threshold ($k = 1.5$ by default).

```
# Applying detect_outliers_loop to the simulated dataset

data_cleaned <- detect_outliers_loop(simulated_data, "tests_HIV", k = 1.5)

# Interactive options for handling outliers: "remove", "keep", or "mark"

# Viewing the dataset after outlier management

head(data_cleaned)
```

When utilising the **detect_outliers_loop** function for outlier detection, you will encounter interactive prompts that offer three choices for handling identified outliers: 'remove', 'keep' or 'mark.' Here's what each option signifies:

- *Remove*: Selecting this option will delete the identified outliers from your dataset, although it may be warranted if you detect clear data entry errors (e.g., impossible values).
- *Keep*: Choosing to 'keep' allows you to retain the outliers in your dataset without any modification. This approach is often advisable if the outliers hold important information or if your chosen model can accommodate high leverage points (e.g., NB or hurdle approaches that handle overdispersion).
- *Mark*: Opting to 'mark' outliers will flag these data points in your dataset, typically by adding a new column indicating whether each observation is an outlier. This offers a balanced strategy, allowing for further investigation or differentiated analysis of flagged observations without permanently altering or removing them from your dataset.

The function's iterative nature enables you to re-evaluate the dataset after each action, ensuring a thorough outlier management process tailored to your analysis needs. That said, we recommend investigating outliers in count data carefully to determine whether they reflect plausible real-world events (e.g., participants legitimately engaging in more frequent behaviour) versus actual data errors.

In our data, the outlier rate was $< 5\%$ of observations, so we chose the 'keep' option, given that none appeared implausible or erroneous.

5.2.3 | Data Suitability Check for the Hurdle Model

In this section, we perform two important diagnostic steps:

1. Global check (check_hurdle_data)

We first examine the entire dataset, including zeros, to identify both zero inflation and global overdispersion.

2. Positive subset overdispersion

We then focus specifically on observations with tests_HIV > 0 to determine whether the positive counts still exhibit overdispersion or can be adequately modelled by a Poisson distribution.

```
# Load necessary packages
library(AER)

# (A) Check global data suitability with check_hurdle_data()
results <- check_hurdle_data(data_cleaned, "tests_HIV")

## Your dataset exhibits significant overdispersion and zero inflation, which may make a Hurdle model appropriate.
##
## Data Summary:
## -------------
## Variable:  tests_HIV
##  Mean:  1.688
##  Variance:  2.159
##  Overdispersion Ratio (variance/mean):  1.279
##  A ratio significantly greater than 1 might indicate the suitability of a Hurdle model over a Poisson model.
##
## P-value for Overdispersion Test:  0
##  A p-value less than 0.05 suggests significant overdispersion.
##
## Proportion of Zeros in  tests_HIV :  0.245
##  A high proportion of zeros could suggest zero inflation, making a Hurdle model more suitable.
##
## Proportion of Zeros in Poisson distribution:  0.185
##  This is the proportion of zeros would be expected if the data followed a Poisson distribution.

# This function call examines both overdispersion and zero inflation
# across the entire dataset (including zeros).

# (B) Evaluate overdispersion specifically among positive counts
pos_data <- subset(data_cleaned, tests_HIV > 0)

pos_disp_test <- dispersiontest(
  glm(tests_HIV ~ 1, family = poisson(), data = pos_data)
)

# Print the dispersion test results for the positive subset
pos_disp_test

##
##  Overdispersion test
##
## data:  glm(tests_HIV ~ 1, family = poisson(), data = pos_data)
## z = -7.3585, p-value = 1
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
##  0.7287743
```

Two conclusions can be drawn from these results:

- *Global overdispersion and zero inflation*: the full dataset exhibits a higher proportion of zeros than the Poisson expectation, and variance exceeds the mean (overdispersion ratio ~ 1.28). This outcome already indicates that a two-part model, such as a Hurdle approach, is desirable.
- *Truncated subset not overdispersed*: when we filtered the data to tests_HIV > 0 only and rechecked overdispersion, we found $z = -7.36$ and $p = 1$, suggesting no evidence of overdispersion (in fact, a dispersion estimate of ~ 0.73 indicates mild *under*dispersion). This suggests that for the positive counts, a truncated Poisson distribution may suffice instead of a truncated NB.

5.3 | Fitting a Hurdle Poisson Model

Now that our data has been thoroughly prepared, we're ready to implement the Hurdle model.

5.3.1 | Loading Necessary Packages

First, ensure that the pscl package (Zeileis et al. 2008), which contains functions for count data models including the Hurdle model, is installed and loaded into your R session.

```
if (!requireNamespace("pscl", quietly = TRUE)) {
  install.packages("pscl")
}
library(pscl)
```

5.3.2 | Fitting the Hurdle Model With Distinct Predictors

Here's how we fit a Hurdle model with separate predictors for each component. We hypothesise that factors such as age, education level and access to healthcare influence the decision to undergo HIV testing (the logistic component), while perceived risk, sexual behaviour risk and social support affect the frequency of testing among those who choose to get tested (the count component).

In this model specification, the | operator effectively separates the predictors for the count component of the model (to the left of |), which is applied to nonzero counts, from those for the logistic regression component (to the right of |), which models the binary outcome of zero vs. nonzero counts. This distinction allows us to investigate how different factors influence the decision to get tested and the testing frequency among those who decide to proceed.

```r
# Fitting the Hurdle model with different predictors for the logistic and count components
hurdle_model_distinct <- hurdle(tests_HIV ~
access_healthcare + sexual_behavior_risk + perceived_risk |
                             perceived_risk + age + education_level  + social_support,
                             data = data_cleaned,
                             dist = "poisson",   # Poisson for the positive counts
                             zero.dist = "binomial",
                             link = "logit") # For zero inflation in binary decision to test


# Displaying the model summary
summary(hurdle_model_distinct)

##
## Call:
## hurdle(formula = tests_HIV ~ access_healthcare + sexual_behavior_risk +
##     perceived_risk | perceived_risk + age + education_level + social_support,
##     data = data_cleaned, dist = "poisson", zero.dist = "binomial", link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.3678 -0.6340 -0.2875  0.6373  4.9663
##
## Count model coefficients (truncated poisson with log link):
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)           0.43118    0.07490   5.757 8.57e-09 ***
## access_healthcare     0.04686    0.05247   0.893   0.3717
## sexual_behavior_risk  0.15781    0.09078   1.738   0.0822 .
## perceived_risk        0.21239    0.09104   2.333   0.0196 *
## Zero hurdle model coefficients (binomial with logit link):
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      0.10693    0.53684   0.199 0.842119
## perceived_risk   0.13816    0.22620   0.611 0.541350
## age              0.02639    0.02048   1.288 0.197613
## education_level -0.20913    0.13011  -1.607 0.107989
## social_support   0.83258    0.22433   3.711 0.000206 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 9
## Log-likelihood: -2176 on 9 Df
```

5.3.2.1 | Interpreting the Hurdle Model Summary.

The Hurdle model summary provides insights into how various factors influence both the decision to engage in an event (e.g., undergoing HIV testing) and the frequency of the event among those who decide to engage. Here's how to interpret significant predictors in the model:

- *Count model coefficients*: These coefficients reflect the impact of predictors on the frequency of the event among those who have crossed the hurdle (i.e., decided to engage in the event). For example, a significant positive coefficient for perceived_risk suggests that as perceived risk increases by 1 unit, the frequency of HIV testing increases, holding other variables constant.
- *Zero Hurdle model coefficients*: In the Hurdle model, the logistic regression component predicts the probability of not engaging in the event at all (i.e., probability of having zero counts for the outcome variable). Thus, a positive coefficient for a predictor in this component indicates an increase in the probability of not adopting the behaviour, or equivalently a decrease in the probability of engaging at least once in the event. For example, if social_support has a significant positive coefficient, it means that higher levels of social support are associated with an increased probability of not getting tested. In other words, more social support is associated with a lower probability of getting tested for HIV at least once. Conversely, a negative coefficient would indicate a higher likelihood of engaging in the event at least once. The zero hurdle model predicts the odds of observing a zero versus a nonzero count using logistic regression.
- *Pearson residuals*: Evaluate the fit of the model. Substantial residuals may indicate misfit or the presence of outliers.
- *Theta* (*count dispersion*): A measure of overdispersion in count data. A significant theta suggests that the NB distribution is appropriate for modelling the count data due to overdispersion.
- *Model fitting and efficiency*: Details like the number of iterations and the log-likelihood value give an idea of how well the model fits the data and the efficiency of the optimisation process.

5.3.2.2 | Example of Reporting Model Results (APA Style).

When examining factors influencing HIV testing frequency among young MSM using a Hurdle Poisson model, we found that perceived risk significantly increased the frequency of testing ($\beta = 0.212$, SE $= 0.091$, $z = 2.333$, $p = 0.02$). Specifically, a 1-unit increase in perceived risk was associated with a 23.6% rise in the expected number of HIV tests ($\exp(0.212) = 1.236$). In contrast, the zero part of the model indicated that social support had a significant positive coefficient ($\beta = 0.833$, SE $= 0.224$, $z = 3.711$, $p < 0.001$), suggesting that higher levels of social support were linked to an increased likelihood of remaining at zero tests—equivalently, a lower probability of undergoing HIV testing at least once.

5.4 | Evaluation of the Hurdle Model

After fitting a Hurdle model to our data, it's crucial to evaluate its performance to ensure it provides a reliable representation of the underlying data generating process. This section will guide you through various methods to assess the quality and fit of our Hurdle model, offering insights into potential improvements.

5.4.1 | Importance of Model Evaluation

Evaluating the Hurdle model is essential for several reasons. It helps us to understand how well the model captures the complexities of our data, including overdispersion and zero-inflation.

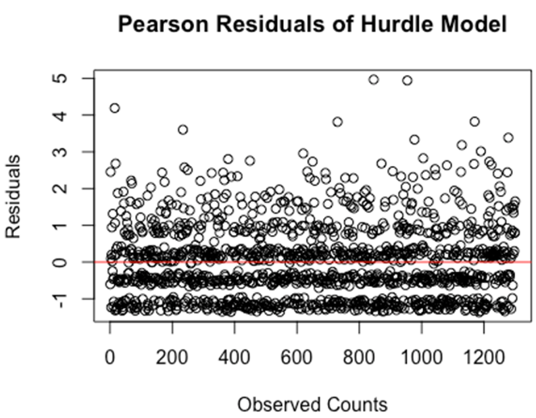
Furthermore, a thorough evaluation can reveal areas where the model may fall short, guiding us towards refinements and adjustments for better predictions.

5.4.2 | Methods for Evaluating the Hurdle Model

5.4.2.1 | Residual Analysis. One fundamental approach is to examine the residuals of the model. Residuals, the differences between observed and predicted values, can indicate how well the model fits the data. For the Hurdle model, we look at the Pearson residuals as a primary diagnostic tool.

```
# Calculating Pearson residuals
residuals <- residuals(hurdle_model_distinct, type = "pearson")

# Plotting residuals to check for patterns
plot(residuals, type = 'p', main = "Pearson Residuals of Hurdle Model",
    xlab = "Observed Counts", ylab = "Residuals")
abline(h = 0, col = "red")
```



The plot of Pearson residuals from the Hurdle model is a diagnostic tool for evaluating model fit. Here's what to look for:

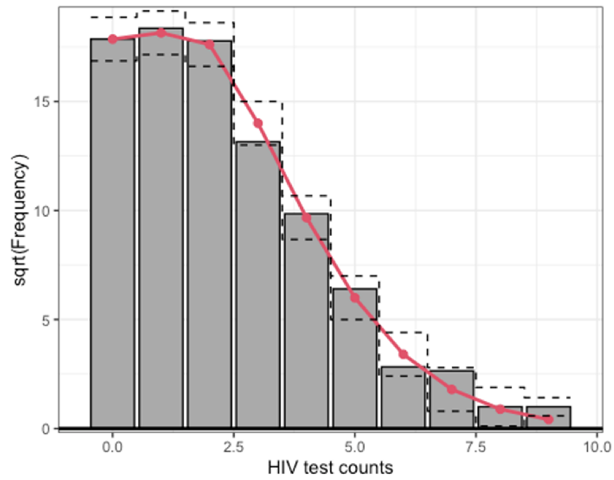
- *Residual scatter*: We expect residuals to be scattered randomly around the horizontal line at zero, which represents no difference between observed and predicted values. Consistent patterns or trends in the residuals, such as a systematic curve or clusters, may indicate that the model is not capturing some aspect of the data's structure.
- *Outliers*: Points far from the zero line suggest the model has not captured these observations well. Occasional outliers are expected, but numerous or extreme outliers may indicate the model is missing key variables or interactions, or that there are data quality issues.
- *Zero line crossing*: Most residuals should cross the zero line evenly. An uneven crossing could hint at model mis-specification.

In our plot, the residuals appear randomly distributed, indicating no obvious systematic errors in the model. Some outliers are present, suggesting possible individual data points that the model fails to predict accurately. However, without clear patterns or a directional trend, the model seems to fit the majority of the data adequately.

5.4.2.2 | Tukey Rootogram. A Tukey rootogram (Tukey 1972) can help assess the fit of count models (Kleiber and Zeileis 2016), including the Hurdle model, by comparing observed and expected frequencies of counts. The topmodels package provides infrastructure for assessment of probabilistic models, including specialized visualization tools like rootograms (Zeileis et al. 2024)

```
# Ensure the 'topmodels' package is installed for rootograms
# This package provides advanced modeling and visualization tools
if (!requireNamespace("topmodels", quietly = TRUE)) {
  install.packages("topmodels", repos = "https://R-Forge.R-project.org")
}
library(topmodels)


# Generating the rootogram for the Hurdle model
rootogram_object <- topmodels::rootogram(hurdle_model_distinct,
                                         plot = TRUE,
                                         xlim = c(-1, 10),
                                         style = "standing",
                                         main = "Rootogram of Hurdle Model",
                                         xlab = "HIV test counts")
```




```
print(rootogram_object)

## A `rootogram` object with `scale = "sqrt"` and `style = "standing"`
## (column `distribution` not shown)
##
##     observed   expected mid width
## 1  17.860571 17.8605710   0   0.9
## 2  18.357560 18.1476415   1   0.9
## 3  17.776389 17.6129779   2   0.9
## 4  13.152946 14.0031994   3   0.9
## 5   9.848858  9.6734993   4   0.9
## 6   6.403124  5.9967194   5   0.9
## 7   2.828427  3.4046585   6   0.9
## 8   2.645751  1.7954198   7   0.9
## 9   1.000000  0.8884821   8   0.9
## 10  1.000000  0.4158306   9   0.9
```

In the rootogram of the hurdle model shown, the grey bars depict the square roots of the observed frequencies for HIV tests, while the red dots represent the square roots of the frequencies predicted by the model. A tight alignment of the red dots with the tops of the grey bars signals a good fit of the model for these values. Here, for counts of zero and one HIV test, the model appears to provide an accurate fit. Its goodness-of-fit seems to diminish for subsequent higher counts. When the red dot lies below the top of the grey bar, the model underestimates the observed frequency; when the red dot is above the grey bar, the model overestimates the observed frequency. However, the discrepancy between observed and estimated frequencies is considered statistically reasonable when the red dot does not extend beyond the horizontal dashed lines, which indicate the theoretical limit of statistical reliability for low frequency counts in the data.

In case of discrepancies or areas of concern, such as patterns in residuals or poor fit indicated by the rootogram, it might be necessary to revisit the model specification. This could include considering alternative distributions for the count data, adjusting the predictors for each component of the model, or exploring other models better suited to the data.

6 | Practical Considerations, Pitfalls and Model Assumptions

Despite its versatility, the hurdle model is not a panacea; researchers must remain attentive to several conceptual and methodological considerations.

6.1 | Model Assumptions

Independence of observations: Each observation should be statistically independent. If your data involve nested structures (e.g., repeated measures within individuals), a mixed-model extension may be necessary[2].

No perfect multicollinearity: Predictors must not be perfectly correlated, as this can inflate standard errors and obscure the distinct contribution of each variable.

Appropriate distribution after separating zeros: Hurdle models split the data into zeros (binary part) and strictly positive counts (count part). For the positive subset, confirm whether a Poisson or NB distribution (or another distribution, e.g., Gamma) is most suitable, based on tests of overdispersion or distribution shape.

No structural zeros: Hurdle models assume that all zeros stem from the same binary 'hurdle' decision. If theory suggests some participants truly cannot produce a positive count—for instance, a structural impossibility—then zero-inflated models (e.g., ZIP, ZINB) may be more appropriate (Loeys et al. 2012).

6.2 | Common Pitfalls

Some real-world complications can arise when using a hurdle model. This section explains frequent pitfalls and offers guidance to keep analyses on track.

Overdispersion considerations: Even after modelling zeros separately, the positive part can remain overdispersed. A NB hurdle (Hilbe 2011) can be used if variance still greatly exceeds the mean. Conversely, if the positive subset is not overdispersed (or is underdispersed), a Poisson hurdle is usually appropriate.

Potential structural zeros: Another pitfall lies in misidentifying zeros as 'sampling' when they are, in fact, 'structural'. If some participants physically cannot produce any positive count, or they are systematically excluded from an opportunity to engage in the behaviour, a zero-inflated approach may be warranted (Lambert 1992). Discerning the nature of zeros is crucial.

Excessive zero inflation: Occasionally, even a hurdle specification fails to capture the sheer number of zeros observed. Researchers should verify whether additional factors (e.g., crucial predictors omitted from the model) might explain these zeros. If the model consistently underestimates zeros, exploring alternative zero-augmented frameworks or refining the zero part of the hurdle (e.g., different link functions or predictor sets) can help.

Predictor selection: All predictors included in both the binary and count components should be grounded in theoretical and empirical rationale. As Mullahy (1986) emphasised, the logic of 'two-part' models relies on the understanding that the decision to engage and the frequency/intensity of engagement may be governed by separate (though related) sets of factors.

Continuous model improvement: It is rare to specify a perfect hurdle model on the first attempt. Instead, the modelling process is iterative, requiring researchers to: (1) revise predictor sets based on diagnostic feedback, theoretical updates, or newly identified confounders, (2) monitor residuals, rootograms and other fit indices to detect areas of misalignment, and (3) compare alternative hurdle specifications (Poisson vs. NB, different covariates, etc.) using criteria like AIC or BIC.

Thus, the development of a robust hurdle model requires an iterative approach that extends beyond the initial data analysis. It requires systematic examination of the underlying data-generating mechanisms, judicious selection of predictor variables and continuous reevaluation of methodological decisions as additional empirical evidence becomes available.

7 | Conclusion

This manuscript has explored the utility of hurdle models as a dual-component framework that distinguishes between the initial participation decision and the subsequent intensity of engagement among active participants. This methodology proves particularly valuable in behavioural research settings, as exemplified by Dr. Efigenia Nasa's HIV testing investigation, where the data exhibit a distinct bifurcation between non-participants and those showing varying degrees of engagement.

7.1 | Key Takeaways

- *Overdispersion handling*: By segregating zeros and positive counts, hurdle models address instances where the overall variance far exceeds the mean.
- *Zero adaptability*: The binary hurdle component directly models excess zeros, often providing more intuitive interpretations than zero-inflated alternatives that assume structural zeros.
- *Interpretive clarity*: A single source of zeros removes the ambiguity about which zeros come from which mechanism, simplifying coefficient interpretation in the zero part of the model.

- *Broad applicability*: Hurdle logic applies not only to count data but to any distribution with a marked point-mass inflation, including continuous or proportion data (via Gamma or Beta hurdles).

We encourage researchers to explore further developments—such as multilevel hurdle models (Brooks et al. 2017) or hurdle approaches for continuous outcomes—and to adopt an iterative, theory-driven mind set. In doing so, we harness the strengths of hurdle models to gain richer, more nuanced insights into the behaviours and phenomena at the heart of psychological inquiry.

Ethics Statement

As this study does not involve human participants, animals or sensitive data, ethical approval was not required. However, the research adheres to the principles outlined in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent

As this study is based on simulated data and does not involve human participants, no informed consent was required.

Conflicts of Interest

The author declares no conflicts of interest.

Endnotes

[1] We focus here on zero inflation, but it is also possible to encounter zero deflation, that is, fewer zeros than expected under standard distributions. In these cases, zero-modified distributions (Conceição et al. 2017) can be advantageous because they allow for either zero-inflation or zero-deflation, extending the logic of hurdle or zero-inflation modelling to situations with unexpected zero frequencies in either direction.

[2] Hierarchical or mixed-model hurdle approaches (e.g., via glmmTMB) can accommodate random effects in both the hurdle and count components. Such extensions relax the independence assumption by modelling subject- or group-level random intercepts (or slopes) for each part of the hurdle framework.

References

Brooks, M. E., K. Kristensen, K. J. Van Benthem, et al. 2017. "glmmTMB Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modeling." *R Journal* 9, no. 2: 378–400. https://doi.org/10.32614/RJ-2017-066.

Conceição, K. S., F. Louzada, M. G. Andrade, and E. S. Helou. 2017. "Zero-Modified Power Series Distribution and Its Hurdle Distribution Version." *Journal of Statistical Computation and Simulation* 87, no. 9: 1842–1862. https://doi.org/10.1080/00949655.2017.1289529.

Coxe, S., S. G. West, and L. S. Aiken. 2009. "The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives." *Journal of Personality Assessment* 91, no. 2: 121–136. https://doi.org/10.1080/00223890802634175.

Cragg, J. G. 1971. "Some Statistical Models for Limited Dependent Variables With Application to the Demand for Durable Goods." *Econometrica: Journal of the Econometric Society* 39: 829–844.

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, no. 1: 8. https://doi.org/10.1186/s40488-021-00121-4.

Greene, W. H. 1994. "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models." Working Paper No. EC-94-10. NYU. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1293115.

Heilbron, D. C. 1994. "Zero-Altered and Other Regression Models for Count Data With Added Zeros." *Biometrical Journal* 36, no. 5: 531–547. https://doi.org/10.1002/bimj.4710360505.

Hilbe, J. M. 2011. *Negative Binomial Regression*. Cambridge University Press. https://books.google.fr/books?hl=fr&lr=&id=0Q_ijxOEBjMC&oi=fnd&pg=PR11&dq=Negative+binomial+regression&ots=IL23WqxXOb&sig=cxQkNbZBTIO6IbY9u6-UhxsDIZE.

Kleiber, C., and A. Zeileis. 2016. "Visualizing Count Data Regressions Using Rootograms." *American Statistician* 70, no. 3: 296–303. https://doi.org/10.1080/00031305.2016.1173590.

Lambert, D. 1992. "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing." *Technometrics* 34, no. 1: 1–14. https://doi.org/10.2307/1269547.

Loeys, T., B. Moerkerke, O. De Smet, and A. Buysse. 2012. "The Analysis of Zero-Inflated Count Data: Beyond Zero-Inflated Poisson Regression." *British Journal of Mathematical and Statistical Psychology* 65, no. 1: 163–180.

Mullahy, J. 1986. "Specification and Testing of Some Modified Count Data Models." *Journal of Econometrics* 33, no. 3: 341–365.

Tukey, J. W. 1972. "Some Graphic and Semi-Graphic Displays." In *Statistical Papers in Honor of George W. Snedecor*, edited by T. A. Bancroft and S. A. Brown, vol. 5, 293–316. Iowa State University Press. http://www.edwardtufte.com/tufte/tukey.

Vives, J., J.-M. Losilla, and M.-F. Rodrigo. 2006. "Count Data in Psychological Applied Research." *Psychological Reports* 98, no. 3: 821–835.

Wang, L. 2010. "IRT–ZIP Modeling for Multivariate Zero-Inflated Count Data." *Journal of Educational and Behavioral Statistics* 35, no. 6: 671–692. https://doi.org/10.3102/1076998610375838.

Zeileis, A., C. Kleiber, and S. Jackman. 2008. "Regression Models for Count Data in R." *Journal of Statistical Software* 27, no. 8: 1–25.

Zeileis, A., M. Lang, and R. Stauffer. 2024. "topmodels: Infrastructure for Forecasting and Assessment of Probabilistic Models." (Version 0.3-0/r1758) [R package]. https://R-Forge.R-project.org/projects/topmodels.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.