

Supporting Information (Appendix H)

Rarefaction and Extrapolation: Making Fair Comparison of Abundance-Sensitive Phylogenetic Diversity among Multiple Assemblages

T. C. HSIEH AND ANNE CHAO

Institute of Statistics, National Tsing Hua University, Hsin-Chu 30043, Taiwan

Appendix H. Rarefaction and extrapolation of abundance-sensitive phylogenetic diversity based on incidence data

Model formulation for incidence data

The model formulation of the rarefaction and extrapolation (R/E) with species diversity (Hill numbers) based on incidence data from multiple sampling units was first discussed in Chao et al. (2014, Appendix H). A similar framework was subsequently developed for Faith's PD in Chao et al. (2015, Appendix S7). In this appendix, we extend the previous R/E model of Hill numbers to include phylogenetic differences among species, and we also generalize the previous R/E model of Faith's PD to incorporate species abundances.

To make the material of this appendix self-contained, we duplicate here the model formulation and some derivation steps from Chao et al. (2015, Appendix S7). Suppose in the focal assemblage there are S species indexed by $1, 2, \dots, S$. For any sampling unit, assume that the i th species has its own unique *incidence* (or *occurrence*) *probability* π_i that is constant for any randomly selected sampling unit. The incidence probability π_i is the probability that species i is detected in a sampling unit. We assume that a rooted ultrametric or non-ultrametric phylogenetic tree of the S species (as tip nodes) can be constructed. As in the main text, we assume that all diversity measures are computed from a fixed reference point that is ancestral to all taxa considered in the study.

Assume that there are B branch segments in the tree, $B \geq S$, for the given reference point on the main trunk. Let L_i denote the length of the i th branch. We expand the set of incidence probabilities $(\pi_1, \pi_2, \dots, \pi_S)$ of the S species (as tip nodes) to a larger set of branch incidence probabilities $\{\lambda_i, i = 1, 2, \dots, B\}$ with $(\pi_1, \pi_2, \dots, \pi_S)$ as the first S elements. Here we define λ_i

as the probability of detecting at least one species descended from branch i in a sampling unit, $i = 1, 2, \dots, B$, and refer to λ_i as the branch/node incidence probability of branch i .

Following the unifying approach of Chao et al. (2014), we develop for the first time the phylogenetic diversity measures based on incidence data. Define a phylogenetic entity as a branch segment with unit-length. Consider the collection in which there are L_i phylogenetic entities, $i = 1, 2, \dots, B$, and the abundance of each of these L_i entities is proportional to branch incidence probability λ_i . The total abundance over all entities in the collection is

$\bar{V} = \sum_{j=1}^B L_j \lambda_j$, which is also the mean branch length (weighted by the branch incidence probability). Therefore, the relative abundance for each of the L_i entities is λ_i / \bar{V} so that $\sum_{i=1}^B L_i (\lambda_i / \bar{V}) = 1$. The incidence-data-based phylogenetic diversity of order q is the Hill number of this collection:

$${}^qPD = {}^qPD(\infty) = \left[\sum_{i=1}^B L_i \left(\frac{\lambda_i}{\bar{V}} \right)^q \right]^{1/(1-q)}. \quad (\text{H.1})$$

$${}^1PD = {}^1PD(\infty) = \lim_{q \rightarrow 1} {}^qPD(\infty) = \exp \left(- \sum_{i=1}^B L_i \frac{\lambda_i}{\bar{V}} \log \frac{\lambda_i}{\bar{V}} \right), \quad q = 1. \quad (\text{H.2})$$

This phylogenetic diversity of order q can be interpreted as the effective number of entities, or effective total branch length because each entity is of unit-length. As with abundance data, the diversity order q determines the measures' emphasis on rare or common branches.

A reference sample of size T

For incidence data, “sample size” means “number of sampling units.” We assume that a reference sample of size T (all T sampling units are standardized) is randomly selected from the study area with replacement. The reference sample includes the incidence data of a set of T sampling units. The underlying data consist of a species-by-sampling-unit incidence matrix $\{W_{ij}; i = 1, 2, \dots, S, j = 1, 2, \dots, T\}$ with S rows and T columns; here $W_{ij} = 1$ if species i is detected in sampling unit j , and $W_{ij} = 0$ otherwise, $i = 1, 2, \dots, S, j = 1, 2, \dots, T$. Under our assumption that the probability of detecting species i in any sampling unit is a constant π_i , $i = 1, 2, \dots, S$, the variable W_{ij} follows a Bernoulli distribution with parameter $\pi_i = P(W_{ij} = 1)$. The row sum of the incidence matrix $Y_i = \sum_{j=1}^T W_{ij}$, denotes the *species incidence frequency* of

species $i, i = 1, 2, \dots, S$.

We expand the incidence matrix $\{W_{ij}; i = 1, 2, \dots, S, j = 1, 2, \dots, T\}$ to a larger $B \times T$ matrix $\{W_{ij}, i = 1, 2, \dots, B, j = 1, 2, \dots, T\}$ by redefining that $W_{ij} = 1$ if at least one species descended from branch i is detected in j th sampling unit, and $W_{ij} = 0$ otherwise. This also expands the set of the observed species incidence frequencies $\{Y_1, Y_2, \dots, Y_S\}$ to a larger set $\{Y_i^*, i = 1, 2, \dots, B\}$, which consists of the row sums of the expanded incidence matrix. We refer to Y_i^* as the sample *branch/node incidence frequency* for branch $i, i = 1, 2, \dots, B$. Suppose that the incidence data for all the sampling units are independent. Then $Y_i^*, i = 1, 2, \dots, B$, follows a binomial distribution:

$$P(Y_i^* = y_i) = \binom{T}{y_i} \lambda_i^{y_i} (1 - \lambda_i)^{T - y_i}, \quad y_i = 0, 1, 2, \dots, T,$$

where λ_i is the branch incidence probability. See Figure S7.1 of Chao et al. (2015) for an example. Define R_k as the sum of branch lengths for those branches with incidence frequency k (or incidence probability k/T) i.e.,

$$R_k = \sum_{i=1}^B L_i I(Y_i^* = k), \quad k = 0, 1, \dots, T, \quad (\text{H.3})$$

where $I(\cdot)$ is an indicator function that equals 1 when true and 0 otherwise. Thus, R_0 represents the total length of branches that are not detected in the observed tree (i.e., the tree spanned by the observed species in the reference sample); R_0 is unknown but $\{R_1, R_2, \dots\}$ can be computed from the reference sample and the observed tree. Here R_1 denotes the total branch length of those uniques, and R_2 denotes the total branch length of those duplicates in the branch incidence frequency set $\{Y_i^*, i = 1, 2, \dots, B\}$. Equivalently, R_1 denotes the total branch length of those uniques in the branch incidence frequency set of the observed tree. A similar equivalence is valid for R_2 . The observed phylogenetic diversity in the reference sample can be expressed as

$${}^qPD_{obs} = \left[\sum_{i=1}^B L_i \left(\frac{Y_i^*}{T} \right)^q \right]^{1/(1-q)} = \left[\sum_{k=1}^T \left(\frac{k}{T} \right)^q \times R_k \right]^{1/(1-q)}, \quad (\text{H.4})$$

where $\bar{V}_{obs} = \sum_{i=1}^B L_i (Y_i^* / T)$ is the observed mean branch length in the reference sample.

Notice that when there are no internal nodes in the phylogenetic tree and all branches are equally distinct with branch lengths of unity (i.e., branch lengths are normalized to unity),

$\bar{V}_{obs} = \sum_{i=1}^B Y_i^* / T$, the average number of incidences per assemblage.

^qPD accumulation curve

Although the model formulation and data structure for incidence data are different from those for abundance data, all derivations are generally parallel to those in the main text and thus most details are omitted. To derive the theoretical formula for the expected phylogenetic diversity as a function of sample size $t = 1, 2, \dots$, we assume a hypothetical sample of t sampling units is taken from the entire assemblage with species incidence probabilities $(\pi_1, \pi_2, \dots, \pi_S)$ and branch incidence probabilities $\{\lambda_i, i = 1, 2, \dots, B\}$. As we defined

$\{Y_i^*, i = 1, 2, \dots, B\}$ and $\{R_k, k = 0, 1, \dots, T\}$ for the reference sample of size T , we can similarly define $\{Y_i^*(t), i = 1, 2, \dots, B\}$ and $\{R_k(t), k = 0, 1, \dots, T\}$ for a hypothetical sample of t sampling units. That is, $Y_i(t)$ denotes the sample incidence frequency of branch i , $i = 1, 2, \dots, B$, in a hypothetical sample of size t ; $R_k(t)$ denotes the sum of branch lengths for those branches with incidence frequency k (or incidence probability k/t) in a hypothetical sample of size t , i.e.,

$$R_k(t) = \sum_{i=1}^B L_i I(Y_i^*(t) = k), k = 0, 1, \dots, t. \quad (\text{H.5})$$

Since $Y_i^*(t), i = 1, 2, \dots, B$, follows a binomial distribution with sample size t and probability λ_i , we obtain the analytic formulation of $E[R_k(t)]$:

$$E[R_k(t)] = \sum_{i=1}^B L_i \binom{t}{k} \lambda_i^k (1 - \lambda_i)^{t-k}. \quad (\text{H.6})$$

Define ${}^qPD(t)$ for any positive integer t as the phylogenetic diversity based on the expected distribution of sample branch incidence frequency distribution in a hypothetical sample of size t . For any positive integer t , it follows from Equation (H.6) that the expected mean branch length is $\sum_{k=1}^t (k/t) \times E[R_k(t)] = \sum_{i=1}^B L_i \lambda_i$, which is \bar{V} defined earlier in the model formulation. Using a similar derivation as that for abundance data, we obtain the following formula for ${}^qPD(t)$:

$${}^qPD(t) = \left[\sum_{k=1}^t \left(\frac{k/t}{\bar{V}} \right)^q \times E[R_k(t)] \right]^{1/(1-q)}, \quad q \neq 1; \quad (\text{H.7})$$

$${}^1PD(t) = \exp \left[- \sum_{k=1}^t \left(\frac{k/t}{\bar{V}} \log \frac{k/t}{\bar{V}} \right) \times E[R_k(t)] \right], \quad q = 1. \quad (\text{H.8})$$

We thus obtain the theoretical formulas that are needed to depict the accumulated diversity ${}^qPD(t)$ as a function of the sampling units t . All formulas for $q = 0$, $q = 1$, $q = 2$ and $q > 2$ are shown in the first column of Table H.1.

qPD rarefaction (for incidence data)

Given a reference sample of T sampling units with branch incidence frequency set $\{Y_i, i = 1, 2, \dots, B\}$, the rarefaction problem is to estimate ${}^qPD(t)$, the expected PD in a hypothetical sample of t sampling units ($t < T$) from the assemblage. In this case, an unbiased estimator of $E[R_k(t)]$ exists for $t < T$:

$$\hat{R}_k(t) = \sum_{\substack{i=1 \\ k \leq Y_i^* \leq T-t+k}}^B L_i \frac{\binom{Y_i^*}{k} \binom{T-Y_i^*}{t-k}}{\binom{T}{t}} = \sum_{k \leq j \leq T-t+k} \frac{\binom{T-j}{t-k}}{\binom{T}{t}} R_j. \quad (\text{H.9})$$

Substituting $\hat{R}_k(t)$ and $\bar{V}_{obs} = \sum_{i=1}^B L_i(Y_i^*/T)$ into Equations (H.7) and (H.8), we then obtain the rarefaction estimator of qPD based on incidence data. All formulas for $q=0$, $q=1$, $q=2$ and $q > 2$ are shown in the second column of Table H.1.

qPD extrapolation (for incidence data)

The extrapolation problem is to predict ${}^qPD(T+t^*)$, the expected PD in a hypothetical sample of $T+t^*$ sampling units ($t^* > 0$) from the assemblage. As with the abundance data (Appendix B), we can apply parallel derivations to obtain the extrapolation formulas for incidence data. Here we only summarize the formulas without giving details. The general extrapolation estimator for the phylogenetic diversity of order q is based on the following formula:

$${}^q\widehat{PD}(T+t^*) = {}^qPD_{obs} + [{}^q\widehat{PD}(\infty) - {}^qPD_{obs}][1 - (1 - {}^q\hat{\mu})^{t^*}], \quad q \geq 0. \quad (\text{H.10})$$

Here the estimator ${}^q\hat{\mu}$ is in terms of the estimated asymptote ${}^q\widehat{PD}(\infty)$ (to be specified below for some special cases) and a diversity estimator ${}^q\widehat{PD}(T-1)$ for rarefied sample of size $T-1$ (in Table H.1):

$${}^q\hat{\mu} = [{}^qPD_{obs} - {}^q\widehat{PD}(T-1)] / [{}^q\widehat{PD}(\infty) - {}^q\widehat{PD}(T-1)], \quad q \geq 0. \quad (\text{H.11})$$

See Equation (B.6) of Appendix B for analogous derivations. For the special case that q is an integer ≥ 2 , an exact extrapolation formula is available and given in later text. All formulas for $q = 0$, $q = 1$, $q = 2$ and $q > 2$ are shown in the third column of Table H.1. Three special cases are discussed below.

(a) For $q = 0$, the extrapolation formula is (see Equation B.4 in Appendix B for parallel derivations)

$$\begin{aligned} {}^0\widehat{PD}(T+t^*) &= {}^0PD_{obs} + [{}^0\widehat{PD}(\infty) - {}^0PD_{obs}][1 - (1 - {}^0\hat{\mu})^{t^*}] \\ &= {}^0PD_{obs} + \hat{R}_0^* \left[1 - \left(1 - \frac{R_1}{T\hat{R}_0^* + R_1} \right)^{t^*} \right], \quad t^* \geq 0, \end{aligned} \quad (\text{H.12})$$

where ${}^0\widehat{PD}(\infty) = {}^0PD_{obs} + \hat{R}_0^*$, and \hat{R}_0^* is the Chao2-PD estimator for the undetected 0PD in the reference sample R_0 , i.e.:

$$\hat{R}_0^* = \begin{cases} \frac{(T-1)}{T} \frac{R_1^2}{2R_2} & \text{if } R_2 > \frac{R_1 Q_2^*}{2Q_1}; \\ \frac{(T-1)}{T} \frac{R_1(Q_1-1)}{2(Q_2^*+1)} & \text{if } R_2 \leq \frac{R_1 Q_2^*}{2Q_1}. \end{cases} \quad (\text{H.13})$$

Here Q_1 denotes the number of uniques (i.e., the number of species that occur exactly in one sampling unit) in the observed species incidence frequency set $\{Y_i, i = 1, 2, \dots, S\}$; it is the same as the number of uniques in the sample incidence frequency set of the observed tree because uniques can only occur in the tip nodes. However, Q_2^* (the number of

duplicates in the sample incidence frequency set of the observed tree) is generally different from the number of duplicates in the observed species incidence frequency set $\{Y_i, i = 1, 2, \dots, S\}$.

(b) For $q = 1$, the extrapolation formula is (see Equation B.7 in Appendix B for parallel derivations)

$${}^1\widehat{PD}(T+t^*) = {}^1PD_{obs} + [{}^1\widehat{PD}(\infty) - {}^1PD_{obs}][1 - (1 - {}^1\hat{\mu})^{t^*}],$$

where

$${}^1\hat{\mu} = [{}^1PD_{obs} - {}^1\widehat{PD}(T-1)] / [{}^1\widehat{PD}(\infty) - {}^1\widehat{PD}(T-1)];$$

$${}^1\widehat{PD}(\infty) = \bar{V}_{obs} \exp(\hat{H}_p / \bar{V}_{obs}); \quad \bar{V}_{obs} = \sum_{i=1}^B L_i(Y_i^* / T);$$

${}^1\widehat{PD}(T-1)$: rarefaction estimator at size $T-1$ (Equation H.8);

$$\hat{H}_p(\infty) = \sum_{k=1}^{T-1} \frac{1}{k} \sum_{1 \leq Y_i^* \leq T-k} L_i \frac{Y_i^*}{T} \frac{\binom{T-Y_i^*}{k}}{\binom{T-1}{k}} + \frac{R_1}{T} (1-A^*)^{-T+1} [-\log A^* - \sum_{r=1}^{T-1} \frac{1}{r} (1-A^*)^r];$$

$$A^* = \begin{cases} 2Q_2^* / [(T-1)Q_1 + 2Q_2^*] & \text{if } Q_2^* > 0, \\ 2 / [(T-1)(Q_1 - 1) + 2] & \text{if } Q_2^* = 0, Q_1 \neq 0, \\ 1 & \text{if } Q_2^* = Q_1 = 0. \end{cases}$$

(c) For an integer $q \geq 2$, the extrapolation formula is (see Appendix B for parallel derivations)

$${}^q\widehat{PD}(T+t^*) = \left[\sum_{i=1}^B \frac{L_i}{\bar{V}_{obs}^q} \times \left(\sum_{j=1}^q \psi(q, j) \frac{(T+t^*)^{(j)}}{(T+t^*)^q} \frac{Y_i^{*(j)}}{T^{(j)}} \right) \right]^{1/(1-q)}, \quad (\text{H.14})$$

where $x^{(j)} = x(x-1)\dots(x-j+1)$ denotes the falling factorial, and $\psi(q, j)$ denotes the Stirling number of the second kind defined by the coefficient in the expansion

$x^q = \sum_{j=1}^q \psi(q, j) x^{(j)}$. When t^* tends to infinity, we obtain the following nearly unbiased estimator for the asymptotic diversity of order $q < T$:

$${}^q\widehat{PD}(\infty) = \left[\sum_{\substack{i=1 \\ Y_i \geq q}}^B \frac{L_i}{\bar{V}_{obs}^q} \frac{Y_i^{*(q)}}{T^{(q)}} \right]^{1/(1-q)}. \quad (\text{H.15})$$

In the special case of $q = 2$, Equation (H.14) reduces to

$${}^2\widehat{PD}(T+t^*) = \frac{1}{\sum_{i=1}^B \frac{L_i}{\bar{V}_{obs}^2} \times \left(\frac{1}{T+t^*} \frac{Y_i^*}{T} + \frac{(T+t^*-1)}{T+t^*} \frac{Y_i^*(Y_i^*-1)}{T(T-1)} \right)}, \quad t^* > 0. \quad (\text{H.16})$$

The unconditional variance estimator for the rarefaction estimator ${}^q\widehat{PD}(t)$ and for the extrapolation estimator ${}^q\widehat{PD}(T+t^*)$ with the associated confidence interval can be computed by a bootstrap method. See Appendix S7 in Chao et al. (2015) for details.

TABLE H.1: (Incidence data) The theoretical formulas and analytic estimators for rarefaction and extrapolation of phylogenetic diversity of order $q = 0$ (first row of equations), $q = 1$ (second row), $q = 2$ (third row) and any integer order $q > 2$ (fourth row), given a reference sample¹ of T sampling units.

Theoretical Formula ² (for all $t > 0$)	Interpolation Estimator ³ (for $t < T$)	Extrapolation Estimator ⁴ (for a sample of size $T + t^*$)
${}^0PD(t) = {}^0PD - E[R_0(t)] = \sum_{k=1}^t E[R_k(t)]$	${}^0\widehat{PD}(t) = \sum_{k=1}^t \hat{R}_k(t)$ $= {}^0PD_{obs} - \sum_{\substack{i=1 \\ 1 \leq Y_i^* \leq T-t}}^B L_i \frac{\binom{T-Y_i^*}{t}}{\binom{T}{t}}$ <p style="text-align: center;">(unbiased)</p>	${}^0\widehat{PD}(T+t^*)$ $= {}^0PD_{obs} + [{}^0\widehat{PD}(\infty) - {}^0PD_{obs}][1 - (1 - {}^0\hat{\mu})^{t^*}]$ $= {}^0PD_{obs} + \hat{R}_0^* \left[1 - \left(1 - \frac{R_1}{T\hat{R}_0^* + R_1} \right)^{t^*} \right]$ <p style="text-align: center;">(reliable if $t^* < T$)</p>
${}^1PD(t)$ $= \exp \left[- \sum_{k=1}^t \left(\frac{k/t}{\bar{V}} \log \frac{k/t}{\bar{V}} \right) \times E[R_k(t)] \right]$	${}^1\widehat{PD}(t)$ $= \exp \left[- \sum_{k=1}^t \left(\frac{k/t}{\bar{V}_{obs}} \log \frac{k/t}{\bar{V}_{obs}} \right) \times \hat{R}_k(t) \right]$ <p style="text-align: center;">(nearly unbiased)</p>	${}^1\widehat{PD}(T+t^*)$ $= {}^1PD_{obs} + [{}^1\widehat{PD}(\infty) - {}^1PD_{obs}][1 - (1 - {}^1\hat{\mu})^{t^*}]$ <p style="text-align: center;">(nearly unbiased)</p>
${}^2PD(t) = \frac{1}{\sum_{k=1}^t \left(\frac{k/t}{\bar{V}} \right)^2 \times E[R_k(t)]}$	${}^2\widehat{PD}(t)$ $= \frac{1}{\sum_{k=1}^t \left(\frac{k/t}{\bar{V}_{obs}} \right)^2 \times \hat{R}_k(t)}$ <p style="text-align: center;">(nearly unbiased)</p>	${}^2\widehat{PD}(T+t^*)$ $= \frac{1}{\sum_{i=1}^B \frac{L_i}{\bar{V}_{obs}^2} \times \left(\frac{1}{T+t^*} \frac{Y_i^*}{T} + \frac{(T+t^*-1)}{T+t^*} \frac{Y_i^*(Y_i^*-1)}{T(T-1)} \right)}$ <p style="text-align: center;">(nearly unbiased)</p>

${}^qPD(t)$ $= \left(\sum_{k=1}^t \left(\frac{k/t}{\bar{V}} \right)^q \times E[R_k(t)] \right)^{\frac{1}{1-q}}$	${}^q\widehat{PD}(t) = \left(\sum_{k=1}^t \left(\frac{k/t}{\bar{V}_{obs}} \right)^q \times \hat{R}_k(t) \right)^{1/(1-q)}$ <p style="text-align: center;">(nearly unbiased)</p>	${}^q\widehat{PD}(T+t^*)$ $= \left[\sum_{i=1}^B \frac{L_i}{\bar{V}_{obs}^q} \times \left(\sum_{j=1}^q \psi(q, j) \frac{(T+t^*)^{(j)}}{(T+t^*)^q} \frac{(Y_i^*)^{(j)}}{T^{(j)}} \right) \right]^{1/(1-q)}$ <p style="text-align: center;">(nearly unbiased)</p>
$C(t) = 1 - \frac{\sum_{i=1}^S \pi_i (1 - \pi_i)^t}{\sum_{i=1}^S \pi_i}$	$\hat{C}(t) = 1 - \sum_{Y_i \geq 1} \frac{Y_i}{U} \frac{\binom{T - Y_i}{t}}{\binom{T - 1}{t}}$ <p style="text-align: center;">(nearly unbiased)</p>	$\hat{C}(T+t^*) = 1 - \frac{Q_1}{U} \left[\frac{(T-1)Q_1}{(T-1)Q_1 + 2Q_2} \right]^{t^*+1}$ <p style="text-align: center;">(reliable for $t^* < T$)</p>

NOTES

¹ For the reference sample, the observed phylogenetic diversity of order q is ${}^qPD_{obs}$ (Equation H.4) and the total number of incidence is

$U = \sum_{i=1}^S Y_i$. See Chao et al. (2014, Table 2) for the sample coverage estimator of the reference sample.

² The term $R_k(t)$ is defined in Equation (H.5)

³ An unbiased estimator $\hat{R}_k(t)$ for $E[R_k(t)]$ is given in Equation (H.9) for $t < T$.

⁴ For $q = 0$ and 1, the formulas for ${}^0\widehat{PD}(\infty)$, ${}^1\widehat{PD}(\infty)$, ${}^0\hat{\mu}$ and ${}^1\hat{\mu}$ are given in the text of this Appendix. For any integer $q \geq 2$, $x^{(j)} = x(x-1)\dots(x-j+1)$ denotes the falling factorial, and $\psi(q, j)$ = the Stirling number of the second kind defined by the coefficient in the expansion $x^q = \sum_{j=1}^q \psi(q, j)x^{(j)}$.

⁵ The last row gives equations for sample completeness as a function of sample size, and the corresponding coverage estimators for rarefied samples and extrapolated samples.

References

- Chao A., Chiu, C.-H., Hsieh, T.C., Davis, T., Nipperess, D.A., Faith, D.P. 2015. Rarefaction and extrapolation of phylogenetic diversity. *Methods in Ecology and Evolution*, 6:380-388.
- Chao A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M. 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84:45-67.