Network Meta-Analysis

The Nyaga Arm-Based Model

What is diagnostic test accuracy network meta-analysis?

Diagnostic test accuracy network meta-analysis combines evidence on multiple diagnostic tests within a single statistical model. The evidence network may contain studies evaluating one test and comparative studies evaluating two or more tests.

Studies evaluating several tests provide information about differences between tests under a shared study context. Because results from different tests evaluated in the same study may be statistically dependent, the arm-based model proposed by Nyaga et al. links these test arms through shared study-level random effects and includes:

- test-specific mean logit sensitivities and specificities;
- correlated study-level random effects shared by all tests evaluated in the same study; and
- additional arm-level random effects for individual study-test combinations.

Nyaga, V. N., Aerts, M., & Arbyn, M. (2018). ANOVA model for network meta-analysis of diagnostic test accuracy data. *Statistical Methods in Medical Research*, 27(6), 1766–1784.

How must the data be structured?

The input data must be in arm-based long format. Each row represents one diagnostic test evaluated in one study and contains a complete 2×2 table:

- true positives (`TP`);
- false positives (`FP`);
- false negatives (`FN`); and
- true negatives (`TN`).

A study evaluating several diagnostic tests therefore contributes several rows with the same study identifier.

Consider a study evaluating tests A, B, and C. The study contributes three rows:

Study	Test	TP	FP	FN	TN
Study 1	A	...	...	...	...
Study 1	B	...	...	...	...
Study 1	C	...	...	...	...

Each study-test combination must occur only once. Duplicate rows should be aggregated to one 2×2 table before fitting the model.

Which model does `fitNyaga()` fit?

For study i and diagnostic test k, let Se_{ik} and Sp_{ik} denote sensitivity and specificity, respectively.

The sensitivity model is

$$\text{logit}(Se_{ik}) = \mu_{A,k} + u_{A,i} + \delta_{A,ik},$$

and the specificity model is

$$\text{logit}(Sp_{ik}) = \mu_{B,k} + u_{B,i} + \delta_{B,ik}.$$

The parameters

$$\mu_{A,k} \quad \text{and} \quad \mu_{B,k}$$

are the mean logit sensitivity and specificity of test k.

The study-level random effects are shared by all diagnostic tests evaluated in study i:

$$\begin{pmatrix} u_{A,i} \\ u_{B,i} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right), \quad \text{with} \quad \Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}.$$

What are the arm-level random effects?

The terms

$$\delta_{A,ik} \quad \text{and} \quad \delta_{B,ik}$$

describe residual heterogeneity for individual study-test arms after accounting for the shared study effects.

The model fitted by `fitNyaga()` assumes

$$\delta_{A,ik} \sim \mathcal{N}(0, \tau_A^2) \quad \text{and} \quad \delta_{B,ik} \sim \mathcal{N}(0, \tau_B^2).$$

The arm-level sensitivity and specificity effects are assumed independent.

How are the observed data modelled?

For study-test arm ik, the diseased and non-diseased sample sizes are

$$n_{1,ik} = TP_{ik} + FN_{ik}$$

and

$$n_{0,ik} = TN_{ik} + FP_{ik}.$$

Conditional on the random effects,

$$TP_{ik} \sim \text{Binomial}\left(n_{1,ik}, Se_{ik}\right),$$

and

$$TN_{ik} \sim \text{Binomial}\left(n_{0,ik}, Sp_{ik}\right).$$

The random effects are integrated from the likelihood using the Laplace approximation implemented in Template Model Builder.

How are summary sensitivity and specificity defined?

For test k, `fitNyaga()` reports the conditional summary sensitivity and specificity

$$Se_k = \text{logit}^{-1}(\mu_{A,k})$$

and

$$Sp_k = \text{logit}^{-1}(\mu_{B,k}).$$

These quantities represent the estimated accuracy of test k when the study-level and arm-level random effects equal zero.

Confidence intervals are first constructed on the logit scale:

$$\hat{\mu}_{A,k} \pm z_{1-\alpha/2} SE(\hat{\mu}_{A,k})$$

and

$$\hat{\mu}_{B,k} \pm z_{1-\alpha/2} SE(\hat{\mu}_{B,k}),$$

and are subsequently transformed using the inverse-logit function.

This reporting convention is consistent with the Reitsma models in **dtametaTMB**.

How are contrasts between diagnostic tests reported?

The first level of the test factor is treated as the reference test.

For each non-reference test k, `fitNyaga()` reports the logit-scale contrasts

$$\nu_{A,k} = \mu_{A,k} - \mu_{A,1} \quad \text{and} \quad \nu_{B,k} = \mu_{B,k} - \mu_{B,1}.$$

A positive value of $\nu_{A,k}$ indicates that test k has a higher mean logit sensitivity than the reference test. A positive value of $\nu_{B,k}$ indicates a higher mean logit specificity.

The covariance between the test-specific estimates is incorporated when calculating the standard errors of these contrasts.

The test factor should be releveled before fitting when a particular reference test is required.

Which additional summary measures are reported?

For each diagnostic test, `fitNyaga()` calculates:

Positive likelihood ratio

$$LR_k^+ = \frac{Se_k}{1 - Sp_k}.$$

Negative likelihood ratio

$$LR_k^- = \frac{1 - Se_k}{Sp_k}.$$

Diagnostic odds ratio

$$DOR_k = \frac{LR_k^+}{LR_k^-}.$$

Because

$$\text{logit}(Se_k) = \mu_{A,k}$$

and

$$\text{logit}(Sp_k) = \mu_{B,k},$$

the log diagnostic odds ratio is

$$\log(DOR_k) = \mu_{A,k} + \mu_{B,k}.$$

Confidence intervals for likelihood ratios and diagnostic odds ratios are constructed on the logarithmic scale using the delta method and the estimated covariance between $\widehat{\mu}_{A,k}$ and $\widehat{\mu}_{B,k}$.

How is the total heterogeneity defined?

For a future study-test arm, the total latent sensitivity variance is

$$\sigma_A^2 + \tau_A^2,$$

and the total latent specificity variance is

$$\sigma_B^2 + \tau_B^2.$$

Because the arm-level sensitivity and specificity effects are independent, their covariance remains

$$\sigma_{AB}.$$

The total predictive covariance matrix is therefore

$$\Omega = \begin{pmatrix} \sigma_A^2 + \tau_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 + \tau_B^2 \end{pmatrix}.$$

The corresponding total latent correlation is

$$\rho_{\text{total}} = \frac{\sigma_{AB}}{\sqrt{(\sigma_A^2 + \tau_A^2)(\sigma_B^2 + \tau_B^2)}}.$$

What are the intra-study correlations?

Two sensitivity estimates from different test arms in the same study share the study-level sensitivity effect $u_{A,i}$. Their latent intra-study correlation is

$$ICC_A = \frac{\sigma_A^2}{\sigma_A^2 + \tau_A^2}.$$

Similarly, the latent intra-study correlation between specificity estimates is

$$ICC_B = \frac{\sigma_B^2}{\sigma_B^2 + \tau_B^2}.$$

Values close to one indicate that most heterogeneity is attributable to the shared study context. Values close to zero indicate that most heterogeneity occurs at the individual study-test-arm level.

How are confidence and prediction regions constructed?

For test k, the confidence region is based on the estimated covariance matrix of

$$\begin{pmatrix} \hat{\mu}_{A,k} \\ \hat{\mu}_{B,k} \end{pmatrix}.$$

The prediction region additionally incorporates the total latent covariance matrix

$$\Omega = \Sigma + \begin{pmatrix} \tau_A^2 & 0 \\ 0 & \tau_B^2 \end{pmatrix}.$$

Prediction regions therefore describe the estimated distribution of the true latent sensitivity and specificity of a future study-test arm rather than the distribution of future observed proportions.

The plotting method uses the finite-sample (F)-based contour convention employed by other plotting methods in **dtametaTMB**.

How are starting values obtained?

`fitNyaga()` obtains initial values from `fitReitsmaSubgroupTMB()` using the diagnostic test as a categorical subgroup variable and a common variance-covariance structure.

For initialization only, the subgroup heterogeneity variances are divided between the study-level and arm-level components.

The HPV dataset

The `HPV` dataset contains 36 study-test arms from 17 studies evaluating:

- conventional cytology (`CC`);
- Hybrid Capture 2 (`HC2`); and
- liquid-based cytology (`LBC`).

Each row represents one observed study-test combination.

```
data("HPV")
head(HPV)
#>                     StudyID Year     Age      Location Test  TP   FP    TN FN
#> 1         Belinson JL 2003 2003 Age >30          Asia  HC2 363 1652  6470 12
#> 2         Belinson JL 2003 2003 Age >30          Asia  LBC 331 1523  6599 44
#> 3            Bigras G 2005 2005 Any age        Europe  HC2  80 1063 12697  2
#> 4            Bigras G 2005 2005 Any age        Europe  LBC  48  445 13315 34
#> 5 Cardenas-Turanzas M 2008 2008 Age >30 North America   CC   7   52   767  9
#> 6 Cardenas-Turanzas M 2008 2008 Age >30 North America  HC2  11   55   764  5
```

How do I select the reference test?

The first factor level is used as the reference category for the reported ν_A and ν_B contrasts.

To use HC2 as the reference test:

```
HPV$Test <- factor(HPV$Test,levels = c("HC2", "CC", "LBC"))
levels(HPV$Test)
#> [1] "HC2" "CC"  "LBC"
```

Changing the reference level changes the reported contrasts but does not change the fitted test-specific mean sensitivities and specificities or the maximized likelihood.

How do I fit the Nyaga model?

```
nyaga <- fitNyaga(data = HPV,
                  TP = TP,
                  FP = FP,
                  FN = FN,
                  TN = TN,
                  study = StudyID,
                  test = Test)
nyaga
#>
#> Nyaga NMA Model
#> ---------------
#>
#> Number of studies   : 17
#> Number of arms      : 36
```

```
#> Number of tests     : 3
#> Optimizer           : Converged
#> Hessian             : Positive definite
#> Max |grad|          : 1.832151e-05
#> -2 log likelihood   : 715.085 ( df = 11 )
#> AIC                 : 737.085
#> BIC                 : 754.503
#>
#> Use summary() for parameter estimates.
summary(nyaga)
#>
#> Summary: Nyaga NMA Model
#> ------------------------
#>
#> Parameter estimates
#> -------------------
#>                  Estimate  Std. Error
#> mu_A.HC2      2.756423100 0.246890120
#> mu_A.CC       0.548963790 0.317998400
#> mu_A.LBC      1.435079300 0.314206820
#> mu_B.HC2      2.122441200 0.152648000
#> mu_B.CC       3.013661600 0.200657310
#> mu_B.LBC      2.453780500 0.192106600
#> sigma2_A      0.082036906 0.220778330
#> sigma2_B      0.193972430 0.109554220
#> sigma_AB     -0.070914053 0.097568005
#> rho_AB       -0.562157060 0.845105110
#> tau2_A        0.669675910 0.284039260
#> tau2_B        0.198520730 0.067797247
#> total_var_A   0.751712820 0.224851850
#> total_var_B   0.392493160 0.107381330
#> icc_A         0.109133310 0.289359310
#> icc_B         0.494205880 0.188261040
#> total_rho_AB -0.130554030 0.173525510
#> nu_A.CC      -2.207459300 0.393499300
#> nu_A.LBC     -1.321343800 0.381989580
#> nu_B.CC       0.891220460 0.201929240
#> nu_B.LBC      0.331339300 0.193267060
#>
#> Sensitivity / specificity
#> -------------------------
#>          type   Estimate conflevel   CI_Lower   CI_Upper
```

```
#> mu_A.HC2 sens 0.94027508      0.95 0.90657606 0.96232367
#> mu_A.CC  sens 0.63389515      0.95 0.48143314 0.76354772
#> mu_A.LBC sens 0.80769149      0.95 0.69407609 0.88603987
#> mu_B.HC2 spec 0.89306529      0.95 0.86095643 0.91846158
#> mu_B.CC  spec 0.95318751      0.95 0.93216177 0.96792081
#> mu_B.LBC spec 0.92083747      0.95 0.88867309 0.94429152
#>
#> Diagnostic odds ratio / likelihood ratios
#> ----------------------------------------
#>               Estimate conflevel     CI_Lower     CI_Upper
#> HC2: DOR 131.481250000      0.95 76.625052000 225.60924000
#> HC2: LR+   8.792982500      0.95  6.742371800  11.46726200
#> HC2: LR-   0.066876322      0.95  0.042533067   0.10515213
#> CC: DOR   35.255637000      0.95 17.480255000  71.10651000
#> CC: LR+   13.541155000      0.95  8.914029600  20.57014400
#> CC: LR-    0.384084820      0.95  0.259113400   0.56933043
#> LBC: DOR  48.855147000      0.95 24.602160000  97.01690500
#> LBC: LR+  10.202952000      0.95  7.161215000  14.53667200
#> LBC: LR-   0.208840880      0.95  0.127296610   0.34262118
#>
#> Recovered HSROC parameters (Rutter-Gatsonis)
#> -------------------------------------------
#>        Lambda        Theta       beta sigma2_alpha sigma2_theta
#> HC2 5.1296540  0.853224210 0.43027341   0.11046456  0.098530194
#> CC  3.1110431 -0.874789210 0.43027341   0.11046456  0.098530194
#> LBC 3.7583485 -0.099630891 0.43027341   0.11046456  0.098530194
#>
#> Tests
#> -----
#> [1] "HC2" "CC"  "LBC"
```

How do I interpret the heterogeneity parameters?

The model separates heterogeneity into two components.

The study-level covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix}.$$

This component represents differences in study contexts that affect all diagnostic tests evaluated in the same study.

The arm-level variances

$$\tau_A^2 \quad \text{and} \quad \tau_B^2$$

represent residual variation between individual study-test arms.

The total heterogeneity of a future test arm is determined by

$$\sigma_A^2 + \tau_A^2$$

for sensitivity and

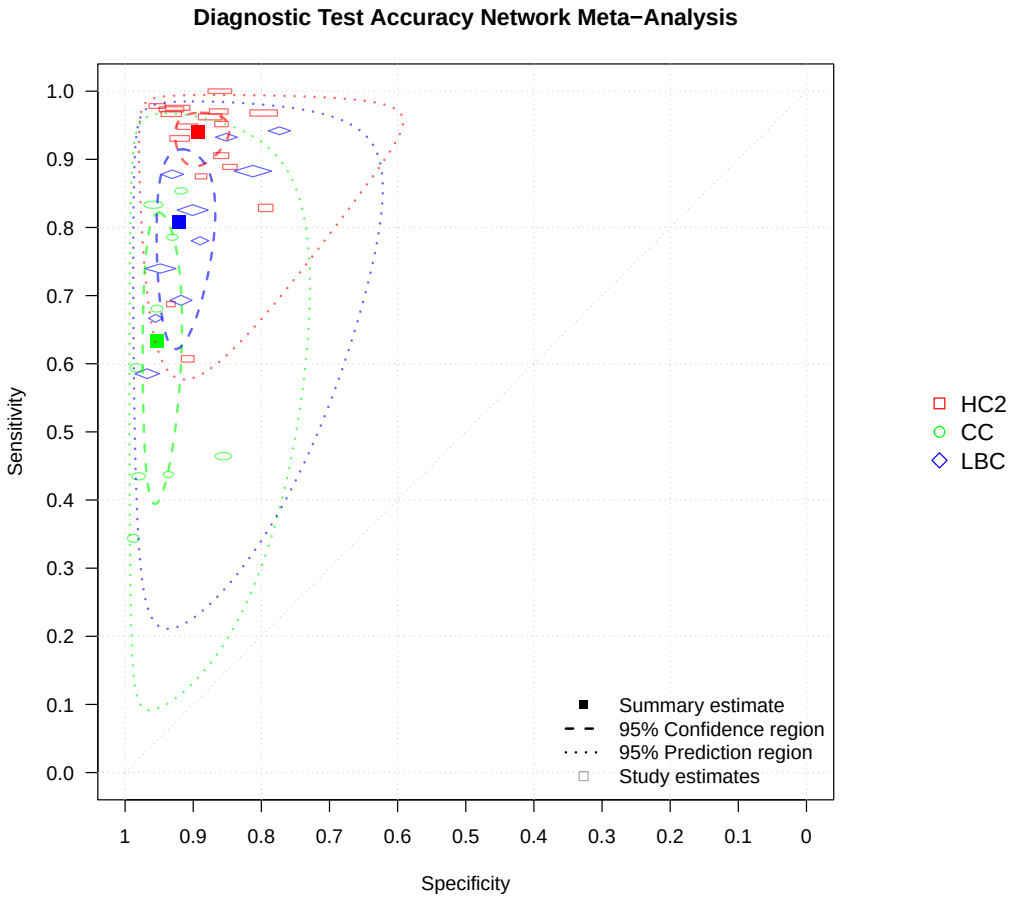
$$\sigma_B^2 + \tau_B^2$$

for specificity.

An arm-level variance may be estimated at or near zero. In such cases, Wald standard errors for that variance component and related quantities may be unreliable, although inference for test-specific fixed effects can remain usable.

How do I obtain a summary ROC plot?

```
plot(nyaga,size="se_revman",nudge_legend = -0.25)
```



How do I obtain a coupled forest plot?

```
forest(nyaga,order = "test")
```

Note: Rendering forest plots may take longer in an interactive R session due to the underlying grid graphics. Performance is typically faster when knitting the vignette (e.g. via RMarkdown or Quarto).

How can I constrain sensitivity across tests?

By default, `fitNyaga()` estimates separate mean logit sensitivities and specificities for every diagnostic test.

To fit a reduced model in which sensitivity is assumed equal across tests:

```
nyaga_common_sens <- fitNyaga(data = HPV,
                              TP = TP,
                              FP = FP,
                              FN = FN,
                              TN = TN,
                              study = StudyID,
                              test = Test,
                              sensspec_constrain = "sens")

nyaga_common_sens
#>
#> Nyaga NMA Model
#> ---------------
#>
#> Number of studies   : 17
#> Number of arms      : 36
#> Number of tests     : 3
#> Optimizer           : Converged
#> Hessian             : Positive definite
#> Max |grad|          : 1.780396e-06
#> -2 log likelihood   : 739.523 ( df = 9 )
#> AIC                 : 757.523
#> BIC                 : 771.775
#>
#> Use summary() for parameter estimates.
```

To assume equal specificity:

```
nyaga_common_spec <- fitNyaga(data = HPV,
                              TP = TP,
                              FP = FP,
                              FN = FN,
                              TN = TN,
                              study = StudyID,
                              test = Test,
                              sensspec_constrain = "spec")

nyaga_common_spec
#>
#> Nyaga NMA Model
```

```
#> ---------------
#>
#> Number of studies   : 17
#> Number of arms      : 36
#> Number of tests     : 3
#> Optimizer           : Converged
#> Hessian             : Positive definite
#> Max |grad|          : 1.471946e-05
#> -2 log likelihood   : 730.315 ( df = 9 )
#> AIC                 : 748.315
#> BIC                 : 762.567
#>
#> Use summary() for parameter estimates.
```

Both may be constrained simultaneously:

```
nyaga_common_both <- fitNyaga(data = HPV,
                              TP = TP,
                              FP = FP,
                              FN = FN,
                              TN = TN,
                              study = StudyID,
                              test = Test,
                              sensspec_constrain = c("sens","spec"))

nyaga_common_both
#>
#> Nyaga NMA Model
#> ---------------
#>
#> Number of studies   : 17
#> Number of arms      : 36
#> Number of tests     : 3
#> Optimizer           : Converged
#> Hessian             : Positive definite
#> Max |grad|          : 1.765095e-05
#> -2 log likelihood   : 754.738 ( df = 7 )
#> AIC                 : 768.738
#> BIC                 : 779.822
#>
#> Use summary() for parameter estimates.
```

How can I compare constrained models?

The constrained models are nested within the unrestricted model and can be compared using likelihood-ratio tests:

```
anova(nyaga_common_sens,nyaga)
#>         Df  logLik Df.diff  Chisq Pr(>Chisq)
#> Model 1  9 -369.76
#> Model 2 11 -357.54       2 24.439  4.935e-06 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(nyaga_common_spec,nyaga)
#>         Df  logLik Df.diff Chisq Pr(>Chisq)
#> Model 1  9 -365.16
#> Model 2 11 -357.54       2 15.23   0.000493 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
anova(nyaga_common_both,nyaga)
#>         Df  logLik Df.diff  Chisq Pr(>Chisq)
#> Model 1  7 -377.37
#> Model 2 11 -357.54       4 39.653  5.106e-08 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

How does `fitNyaga()` differ from `fitReitsmaSubgroup()`?

`fitReitsmaSubgroup()`/`fitReitsmaSubgroupTMB()` can use diagnostic test as a categorical subgroup variable and estimate different mean sensitivities and specificities across tests.

However, a conventional subgroup model does not explicitly represent the dependence created when multiple tests are evaluated within the same study.

`fitNyaga()` introduces a shared study-level random effect:

$$\begin{pmatrix} u_{A,i} \\ u_{B,i} \end{pmatrix},$$

which is common to all tests evaluated in study i. It also includes residual study-test-arm effects:

$$\delta_{A,ik} \quad \text{and} \quad \delta_{B,ik}.$$

The Nyaga model is therefore particularly relevant when the evidence contains a mixture of:

- studies evaluating a single diagnostic test;
- studies comparing two tests; and
- studies evaluating several tests.

A Reitsma subgroup or meta-regression analysis remains useful as a conventional analysis of test differences, especially when each study contributes one primary test result. The models answer related but not identical questions.

What are the current model limitations?

The `fitNyaga()` implementation assumes:

- one common unstructured study-level covariance matrix across tests;
- separate but test-common arm-level sensitivity and specificity variances;
- independent sensitivity and specificity arm-level deviations;
- one 2×2 table per study-test combination; and
- exchangeability of the study-level effects

The current implementation does not estimate:

- test-specific study-level covariance matrices;
- test-specific arm-level variances;
- correlation between sensitivity and specificity arm-level effects;
- additional categorical or continuous covariates;
- network inconsistency parameters; or
- latent-class extensions for an imperfect reference standard.

The model is based on aggregate 2×2 tables and does not distinguish between paired multi-test designs and studies in which separate participant groups receive different index tests. Shared study-level random effects represent dependence arising from a common study context, but the conditional binomial likelihood does not model within-person association between paired test results. A full paired-data likelihood would require joint cross-classifications of the index tests within disease strata, which cannot generally be reconstructed from their marginal 2×2 tables.

How was the implementation validated?

The TMB implementation was compared with the original Bayesian Stan implementation using the Nyaga demonstration data from https://github.com/VNyaga/NMADAS/tree/master/data. The test-specific logit means and study-level covariance estimates were closely comparable. Differences were concentrated in arm-level variance components, particularly where the

maximum-likelihood estimate occurred near the boundary and the Bayesian prior retained a positive posterior distribution.

The HPV data were also fitted using the arm-based network model in Stata `metadta`. The test-specific logit sensitivity and specificity estimates and their standard errors were closely comparable with `fitNyaga()`. However, the Stata fit estimated the study-level correlation at the boundary of -1, and Stata reported that the required curvature condition for the maximized log likelihood was not satisfied.

References

Nyaga, V. N., Aerts, M., & Arbyn, M. (2018). ANOVA model for network meta-analysis of diagnostic test accuracy data. *Statistical Methods in Medical Research*, 27(6), 1766-1784.

Nyaga, V. N., & Arbyn, M. (2023). Comparison and validation of *metadta* for meta-analysis of diagnostic test accuracy studies. *Research Synthesis Methods*, 14(3), 544-562.

Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J. P. M., Bossuyt, P. M., & Zwinderman, A. H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*, 58(10), 982-990.