

# viperData

January 10, 2019

## R topics documented:

dmistData . . . . .	1
dmistRSsummary . . . . .	2
doViperStudyDesign . . . . .	3
doViperSummary . . . . .	3
viperObservations . . . . .	4
viperReaderQualifications . . . . .	6
viperSummary . . . . .	7
<b>Index</b>	<b>9</b>

---

dmistData	<i>DMIST data</i>
-----------	-------------------

---

## Description

DMIST performance data manually transcribed from Pisano2005\_NEJM\_v353p1773 and its supplement.

## Usage

```
dmistData
```

## Format

An object of class `list` of length 6.

## Details

The performance data is organized in a list with the following elements.

- `nobs.FFDM [num]` The number of FFDM observations (from Supplemental Table 2)
- `nobs.SFM [num]` The number of SFM observations (from Supplemental Table 2)
- `nobs.total [num]` the total number of observations (from Supplemental Table 2)
- `caseDist [list]` This list contains two items (from Supplemental Table 2)
  - `desc [char]` A description of the data.

- `caseDist.df` [data frame] The distribution of BIRADS scores from DMIST.
  - Row 1 is the distribution of FFDM BIRADS scores among DMIST cancers.
  - Row 2 is the distribution of SFM BIRADS scores among DMIST cancers.
  - Row 3 is the distribution of FFDM BIRADS scores among all DMIST cases.
  - Row 4 is the distribution of SFM BIRADS scores among all DMIST cases.
- `caseDist.denseBreasts` [data frame] Caption to Figure 1 and row "T stage" of Table 2. One row of cancer=n1 detected by FFDM and SFM, cancer=n1 detected by SFM but not FFDM, cancer=n1 detected by FFDM but not SFM, cancer=n1 not detected by SFM or FFDM, non-cancer=n2, and total.
- `performance` [list]
  - `desc` [char] A description of the data taken from Figure 1 and the text of Pisano2005\_NEJM\_v353p1773.pdf. SE of AUC is taken to be 1/4 of the confidence interval specified. AUC determined using seven-point malignancy scale. TPF,FPF are from Table 2 of Pisano2005\_NEJM\_v353p1773-Suppl.pdf. TPF,FPF are determined using BIRADS. Cancer determined at 455 days for AUC, TPF, and FPF.
  - `FFDM` [list] A list of the diagnostic performance statistics: AUC, AUCse, CIbot, CItop, TPF, TPFse, TNF, TNFse.
  - `SFM` [list] A list of the diagnostic performance statistics: AUC, AUCse, CIbot, CItop, TPF, TPFse, TNF, TNFse.
  - `FFDMminusSFM` [list] A list of the diagnostic performance statistics: AUC, AUCse, CIbot, CItop, p, TPF, TPFse, TNF, TNFse.

dmistRSsummary

dmistRSsummary

## Description

Performance summaries from the following DMIST reader studies:

- Hendrick2008\_Radiology\_v247p38, 'Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACIN DMIST retrospective multireader study.' There were four machines studied, labeled below as hendrick2008.GE, hendrick2008.Fuji, hendrick2008.Fischer, hendrick2008.Hologic. The summary based on the aggregate of all this data is hendrick2008.
- Nishikawa2009\_Radiology\_v251p41, 'Comparison of Soft-Copy and Hard-Copy Reading for Full-Field Digital Mammography.' There were four machines studied, labeled below as nishikawa2009.GE, nishikawa2009.Fuji, nishikawa2009.Fischer, nishikawa2009.Hologic. The summary based on the aggregate of all this data is nishikawa2009.
- An unpublished study evaluating the effect of breast density on diagnostic accuracy of FFDM and SFM. The study included 320 women (109 cancers) stratified by original SFM breast density determination, 8 readers reading FFDM, and 8 readers reading SFM. There were four density levels studied, corresponding to the four BIRADS density levels, labeled as breastDensity.A, breastDensity.B, breastDensity.C, breastDensity.D. The summary based on the aggregate of all this data is breastDensity.
- An unpublished study evaluating performance in a reader study at 50 30 readers assigned to read a total of 300 cases as FFDM and SFM. There was one data set, labeled as prevalence50.

## Usage

```
dmistRSsummary
```

## Format

An object of class `list` of length 16.

---

doViperStudyDesign	<i>doViperStudyDesign</i>
--------------------	---------------------------

---

**Description**

Extract basic study design info

**Usage**

```
doViperStudyDesign(rawData)
```

**Arguments**

`rawData` One of the VIPER sub-study data frames.

**Value**

A list object that contains several ways to summarize the distribution of cases in the study according to case type.

- `nObs` [num] The number of observations in the sub-study.
- `readers` [char] The reader IDs in the sub-study.
- `nR` [num] The number of readers in the sub-study.
- `nC` [data frame] The number of cases in the sub-study by case type. One row of `birads0ffdm`, `birads0sfm`, `birads12ffdm`, `birads12sfm`, `cancer`, `noncancer`.
- `perReader` [list] This list contains two data frames: FFDM and SFM. Each of these data frames includes the number of cases evaluated by each reader by case type. 20 readers == 20 Rows of `birads0ffdm`, `birads0sfm`, `birads12ffdm`, `birads12sfm`, `cancer`, `noncancer`.
- `perReader.summary` [list] This list contains two data frames: FFDM and SFM. Each of these data frames includes the total number of cases evaluated by all readers by case type. One row of `birads0ffdm`, `birads0sfm`, `birads12ffdm`, `birads12sfm`, `cancer`, `noncancer`.

**Examples**

```
df <- split(viperData::viperObs455, viperData::viperObs455$desc)
df.studyDesign <- doViperStudyDesign(df$screeningLowP)
```

---

doViperSummary	<i>doViperSummary</i>
----------------	-----------------------

---

**Description**

This function estimates the VIPER performance results. Before archiving this on GitHub, the function was called `analyzeData`.

**Usage**

```
doViperSummary(viperObs0)
```

**Arguments**

viperObs0      A VIPER raw-data file: [viperObs365](#), [viperObs455](#)

**Value**

A VIPER summary file: [viperSummary365](#) or [viperSummary455](#)

**Examples**

```
library("iMRMC")
viperSummary <- viperData::doViperSummary(viperData::viperObs455)
```

---

viperObservations	<i>viperObservations</i>
-------------------	--------------------------

---

**Description**

The VIPER data sets were created by aggregating all the raw reader data files from the FDA VIPER study: Validation of Imaging Premarket Evaluation and Regulation (VIPER). The data sets differ in the truth labels. See the documentation details.

**Usage**

viperObs365

viperObs455

**Format**

An object of class `data.frame` with 20382 rows and 35 columns.

**Details**

A data frame with 36 variables:

- CaseReadOrder
- CaseID
- Modality
- ReaderID
- Study
- Track
- ReadSetLetter

- Ctype
- Compare.CaseIDs
- Initial.Comparison
- Final.Comparison
- case\_id
- redcap\_event\_name
- readerID
- form\_start\_datetime
- modality
- study
- viewbox
- num\_img
- recall
- miss\_img
- score\_start\_datetime
- abnormality\_type
- recall\_scale
- score\_end\_datetime
- Group
- norecall\_scale
- set\_letter
- track
- quality

- UID
- score
- desc
- read\_order
- Comment
- RecallRegion
- count

**viperObs365:** This dataset is identical to [viperObs455](#) except for the truth labels. The truth labels for this dataset are based on cancer at 365 days.

**viperObs455:** This dataset is identical to [viperObs365](#) except for the truth labels. The truth labels for this dataset are based on cancer at 455 days.

The images used in this study were selected from the DMIST images [Pisano2005\_NEJM\_v353p1773]. The VIPER study concept 2011. Data collection September 2013 to August 2015. Please refer to journal paper for details.

---

vipedReaderQualifications

*VIPER reader qualification summary*

---

## Description

This data set contains the information from the VIPER reader recruitment qualification forms. The original data was modified to turn text in numeric fields to appropriate numbers. For example, "~100" was modified to 100, "100+" was modified to 100, "<50" was modified to 50, "200-250" was modified to 225, "don't know" was modified to "NA", and "very few" was modified to "NA".

## Usage

vipedReaderQualifications

## Format

An object of class `data.frame` with 43 rows and 10 columns.

## Details

A data frame with 10 variables:

- readerID
- certifiedABR

- breastImagingFellowship
- fellowshipYear
- min50sfm
- min50ffdm
- howManySFMlast2years
- howManyFFDMlast2years
- howManyYearsPostResidencyExperienceInterpretingSFM
- howManyYearsPostResidencyExperienceInterpretingFFDM

---

viperSummary

viperSummaries

---

## Description

The VIPER summary files contain summary statistics of the VIPER observations. The iMRMC R package is used to estimate the area under the ROC curve (AUC), the true positive fraction (sensitivity), and the true negative fraction (specificity). The iMRMC package also estimates variances, standard errors, and confidence intervals that account for the variability from readers and cases. You can refer to the R function `doViperSummary.R` in this package for the code that produces the summary files.

The iMRMC R package can be found at

- <https://github.com/DIDSR/iMRMC/releases>,
- <https://github.com/DIDSR/iMRMC/tree/master/Rpackage>, and
- <https://cran.r-project.org/web/packages/iMRMC/index.html>.

## Usage

```
viperSummary365
```

```
viperSummary455
```

## Format

An object of class `list` of length 5.

## Details

Each element of the list pertains to one of the VIPER sub-studies:

- screeningLowP [list] Screening population, per reader prevalence ~ 10%
- screeningMedP [list] Screening population, per reader prevalence ~ 30%
- screeningHighP [list] Screening population, per reader prevalence ~ 50%
- challengeMedP [list] Challenge population, per reader prevalence ~ 30%
- challengeHighP [list] Challenge population, per reader prevalence ~ 50%

Each substudy list element is a list object of length 3.

- desc [string] Sub-study label. This is the same as the sub-study list element name.
- studyDesign [list] This list object contains several ways to summarize the distribution of cases in the study according to case type.
  - nObs [num] The number of observations in the sub-study.
  - readers [char] The reader IDs in the sub-study.
  - nR [num] The number of readers in the sub-study.
  - nC [data frame] The number of cases in the sub-study by case type. One row of birads0ffdm, birads0sfm, birads12ffdm, birads12sfm, cancer, noncancer.
  - perReader [list] This list contains two data frames: FFDM and SFM. Each of these data frames includes the number of cases evaluated by each reader by case type. 20 readers == 20 Rows of birads0ffdm, birads0sfm, birads12ffdm, birads12sfm, cancer, noncancer.
  - perReader.summary [list] This list contains two data frames: FFDM and SFM. Each of these data frames includes the total number of cases evaluated by all readers by case type. One row of birads0ffdm, birads0sfm, birads12ffdm, birads12sfm, cancer, noncancer.
- iMRMC [list] This list object contains the MRMC analysis results for auc, tpf, and tnf. There are four groups of results: perReader, Ustat, MLEstat, ROC, and varDecomp. The analysis results are produced by a command line version of iMRMC called by R: <https://github.com/DIDSR/iMRMC/releases>, <https://github.com/DIDSR/iMRMC/tree/master/Rpackage>, <https://cran.r-project.org/web/packages/iMRMC/index.html>

**viperSummary365:** This summary is based on [viperObs365](#) the dataset based on cancer at 365 days.

**viperSummary455:** This summary is based on [viperObs455](#) the dataset based on cancer at 455 days.



# Index

## \*Topic **datasets**

- dmistData, [1](#)
- dmistRSsummary, [2](#)
- viperObservations, [4](#)
- viperReaderQualifications, [6](#)
- viperSummary, [7](#)

- dmistData, [1](#)
- dmistRSsummary, [2](#)
- doViperStudyDesign, [3](#)
- doViperSummary, [3](#)

- viperObs365, [4](#), [6](#), [8](#)
- viperObs365 (viperObservations), [4](#)
- viperObs455, [4](#), [6](#), [8](#)
- viperObs455 (viperObservations), [4](#)
- viperObservations, [4](#)
- viperReaderQualifications, [6](#)
- viperSummary, [7](#)
- viperSummary365, [4](#)
- viperSummary365 (viperSummary), [7](#)
- viperSummary455, [4](#)
- viperSummary455 (viperSummary), [7](#)