# Validation of open-source packages from CRAN, challenges, process and implementation into different SCE's.

Mark Penniston, Atorus Research

PHUSE SDE 7Sep2023

atorus

Your Complete Data Solution

# Code of Federal Regulations

- Code of Federal Regulations (CFR) is the codification of the general and permanent regulations promulgated by the executive departments and agencies of the Federal government of the United States (50 titles)
  - Title 21:Food and Drugs
    - Chapter 1: Food and Drug Administration
    - Chapter 11: Electronic Records and Signatures


atorus
Your Complete Data Solution

https://www.fda.gov/search?s=software+validation&sort_bef_combine=rel_DESC

An official website of the United States government Here's how you know

U.S. FOOD & DRUG ADMINISTRATION

Search Menu

# Search

software validation

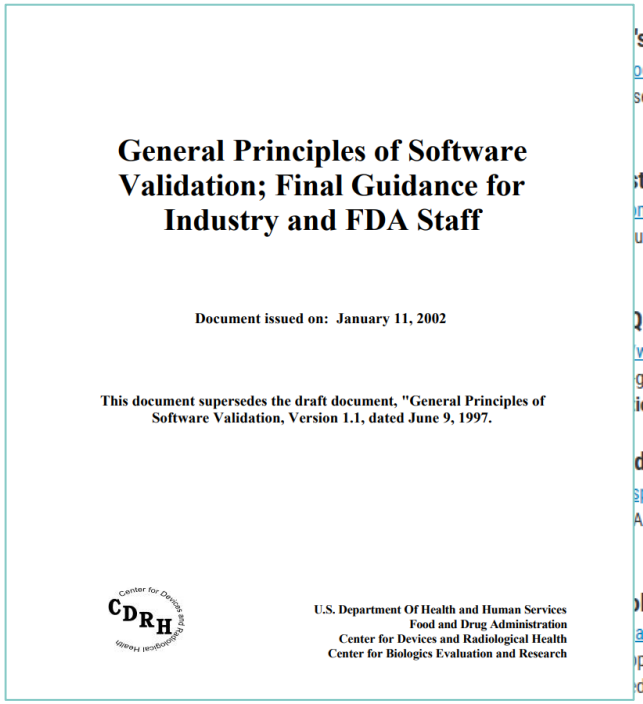
Showing 1 - 25 of 1433 entries

Advanced Search

## General Principles of Software Validation

https://www.fda.gov/regulatory-information/search-fda-guidance-documents/general-principles-software-validation

... nformation/search-fda-guidance-documents/general-principles-**software-validation** ... This guidance outlines general **validation** ... to be applicable to the **validation** of medical device **software** or the **validation** of **software** used to design, ...

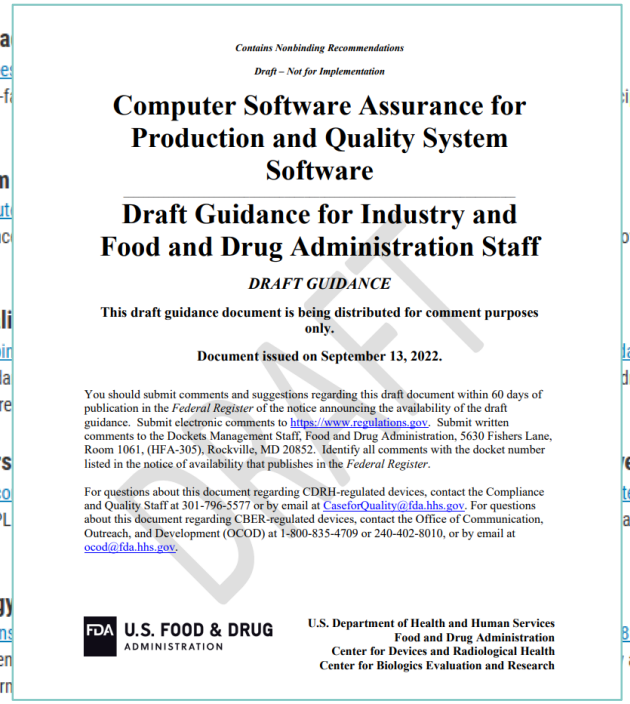


General Principles of Software Validation; Final Guidance for Industry and FDA Staff

Document issued on: January 11, 2002

This document supersedes the draft document, "General Principles of Software Validation, Version 1.1, dated June 9, 1997.

U.S. Department Of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Center for Biologics Evaluation and Research



Contains Nonbinding Recommendations

Draft – Not for Implementation

Computer Software Assurance for Production and Quality System Software

Draft Guidance for Industry and Food and Drug Administration Staff

DRAFT GUIDANCE

This draft guidance document is being distributed for comment purposes only.

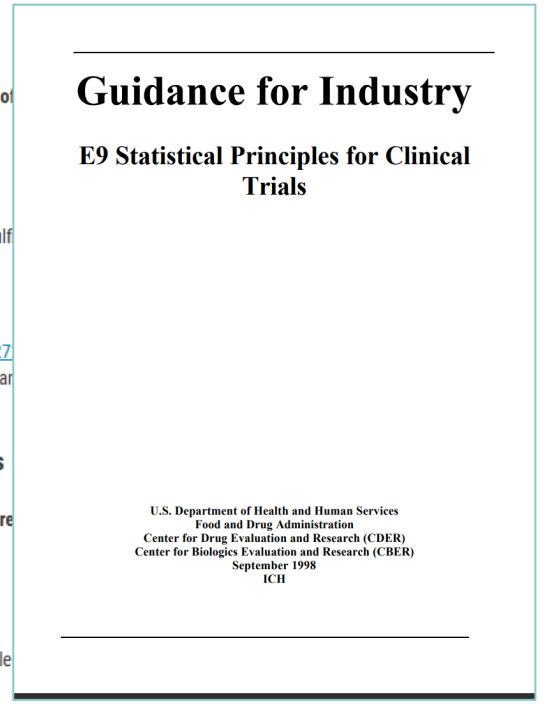
Document issued on September 13, 2022.

You should submit comments and suggestions regarding this draft document within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff, Food and Drug Administration, 5630 Fishers Lane, Room 1061, (HFA-305), Rockville, MD 20852. Identify all comments with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions about this document regarding CDRH-regulated devices, contact the Compliance and Quality Staff at 301-796-5577 or by email at CaseforQuality@fda.hhs.gov. For questions about this document regarding CBER-regulated devices, contact the Office of Communication, Outreach, and Development (OCOD) at 1-800-835-4709 or 240-402-8010, or by email at ocod@fda.hhs.gov.

U.S. FOOD & DRUG ADMINISTRATION

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health
Center for Biologics Evaluation and Research



Guidance for Industry

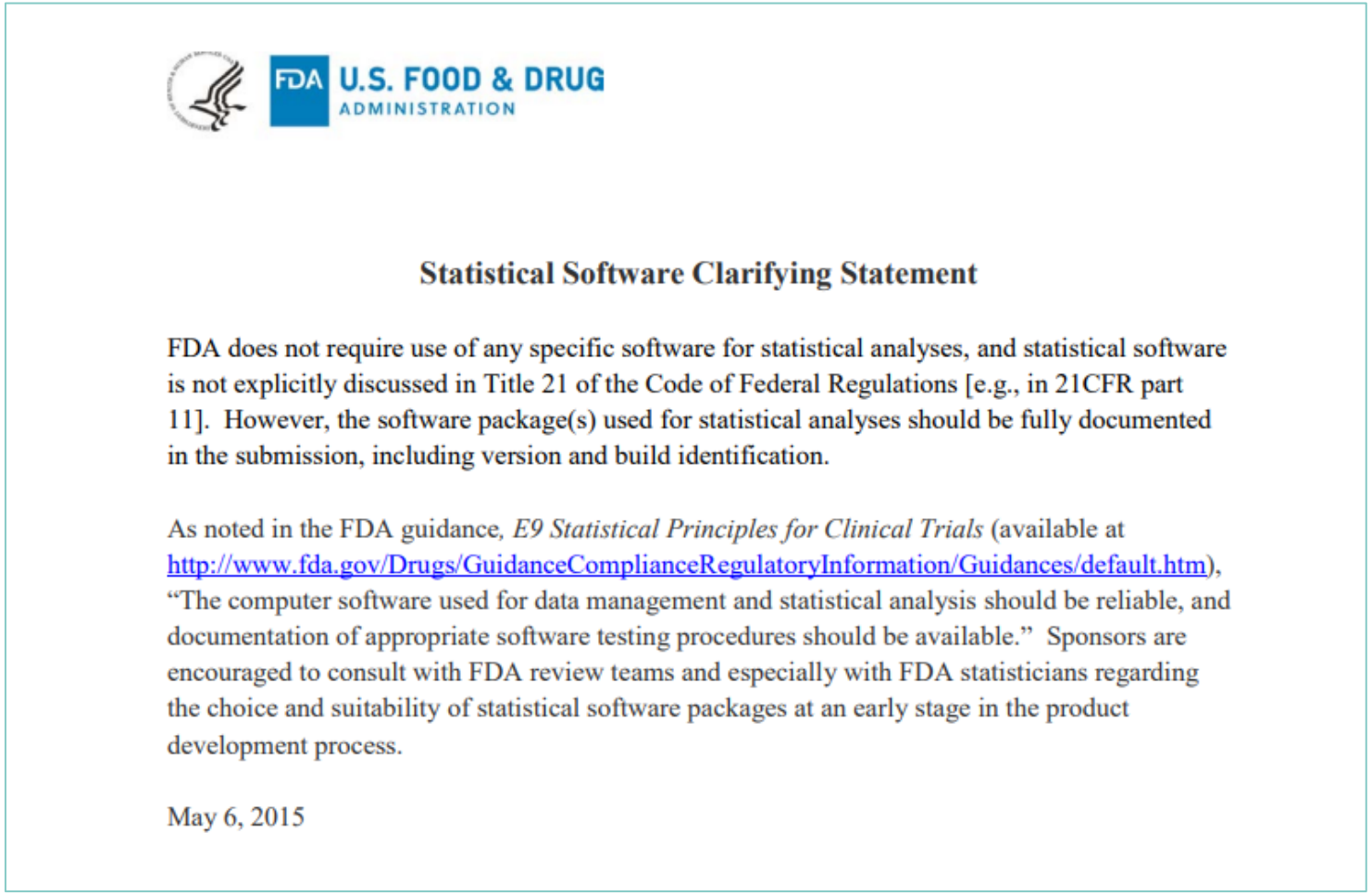
E9 Statistical Principles for Clinical Trials

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
September 1998
ICH

## Step 6: Is the Software Function Intended to Provide Clinical Decision Support?

https://www.fda.gov/medical-devices/digital-health-center-excellence/step-6-software-function-intended-provide-clinical-decision-support

# FDA Position



FDA U.S. FOOD & DRUG ADMINISTRATION

**Statistical Software Clarifying Statement**

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.

As noted in the FDA guidance, *E9 Statistical Principles for Clinical Trials* (available at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm), "The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available." Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process.

May 6, 2015

atorus
Your Complete Data Solution

Base SAS® 9.3 Procedures Guide :: Base SAS(R) 9.3 Procedures Guide, Second Edition

https://support.sas.com/documentation/cdl/en/proc/65145/HTML/default/viewer.htm#titlepage.htm

Sign in | Create Profile

sas THE POWER TO KNOW.

Search

Home Support Learn Connect

Go to Documentation Home

Base SAS(R) 9.3 Procedures Guide, Second Edition

PDF | Purchase

Contents About

Search this document Search

Title Page

What's New in Base SAS 9.3 Procedures

About This Book

Concepts

Procedures

Appendixes

Recommended Reading

Previous Page | Next Page

Base SAS® 9.3 Procedures Guide

Second Edition

Contains the complete reference for all Base SAS procedures. Provides information about what e... pro... u... ...leva... he kin... outp... ...s

Previous Page | Next Page | Top of Page

SAS Support

# July 1, 1976

# SAS BASE – 89 Procedures

# SAS BASE – 1,567 Functions

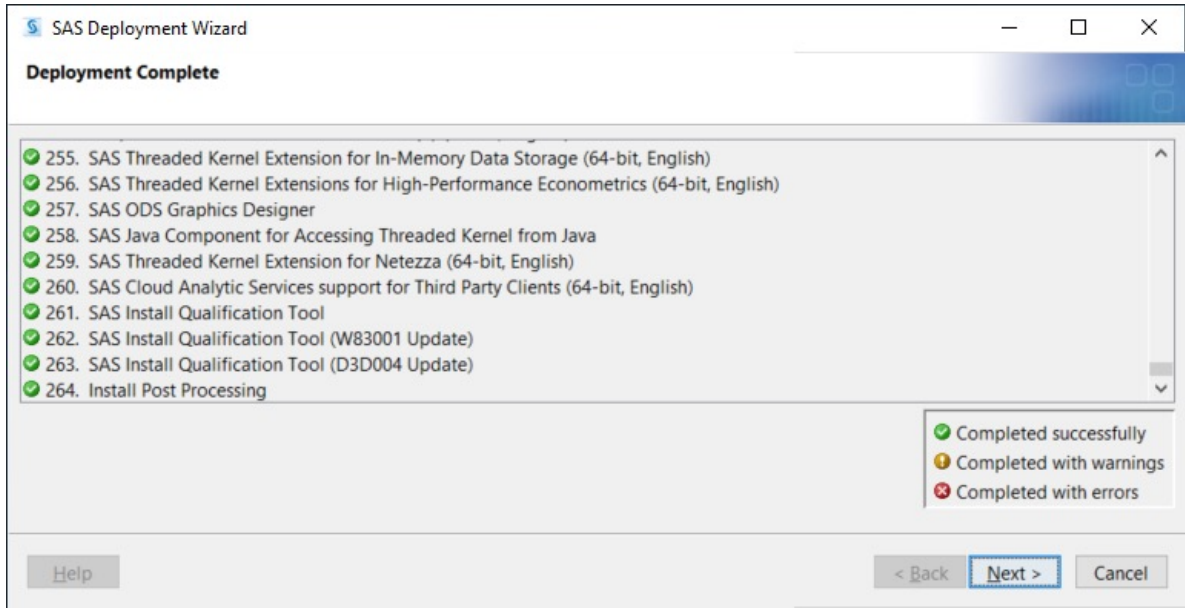
# SAS STAT – 104 Procedures

atorus

Your Complete Data Solution

# Installation & Validation

## Installation




## Validation


- Applies to SAS Macros referenced across studies
- No Risk Assessment

atorus
Your Complete Data Solution

*CRAN*
Mirrors
What's new?
Search
CRAN Team

*About R*
R Homepage
The R Journal

*Software*
R Sources
R Binaries
Packages
Task Views
Other

*Documentation*
Manuals
FAQs
Contributed

# The Comprehensive R Archive Network

## Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux (Debian, Fedora/Redhat, Ubuntu)
- Download R for macOS
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

## Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2023-06-16, Beagle Scouts) R-4.3.1.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are available here. Please read about new features and bug fixes before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages

## Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

## What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. Please consult the R project homepage for further information.

CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN mirror nearest to you to minimize network load.

## Submitting to CRAN

To "submit" a package to CRAN, check that your submission meets the CRAN Repository Policy and then use the web form.

If this fails, send an email to CRAN-submissions@R-project.org following the policy. Please do not attach submissions to emails, because this will clutter up the mailboxes of half a dozen people.

Note that we generally do not accept submissions of precompiled binaries due to security reasons. All binary distribution listed above are compiled by selected maintainers, who are in charge for all binaries of their platform, respectively.

---

*CRAN*
Mirrors
What's new?
Search
CRAN Team

*About R*
R Homepage
The R Journal

*Software*
R Sources
R Binaries
Packages
Task Views
Other

*Documentation*
Manuals
FAQs
Contributed

# Contributed Packages

## Available Packages

Currently, the CRAN package repository features 19783 available packages.

Table of available packages, sorted by date of publication

Table of available packages, sorted by name

CRAN Task Views aim to provide some guidance which packages on CRAN are relevant for tasks related to a certain topic. They provide tools to automatically install all packages from each view. Currently, 43 views are available.

## Installation of Packages

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual R Installation and Administration (also contained in the R base sources) explains the process in detail.

## Package Check Results

All packages are tested regularly on machines running Debian GNU/Linux, Fedora, macOS (formerly OS X) and Windows.

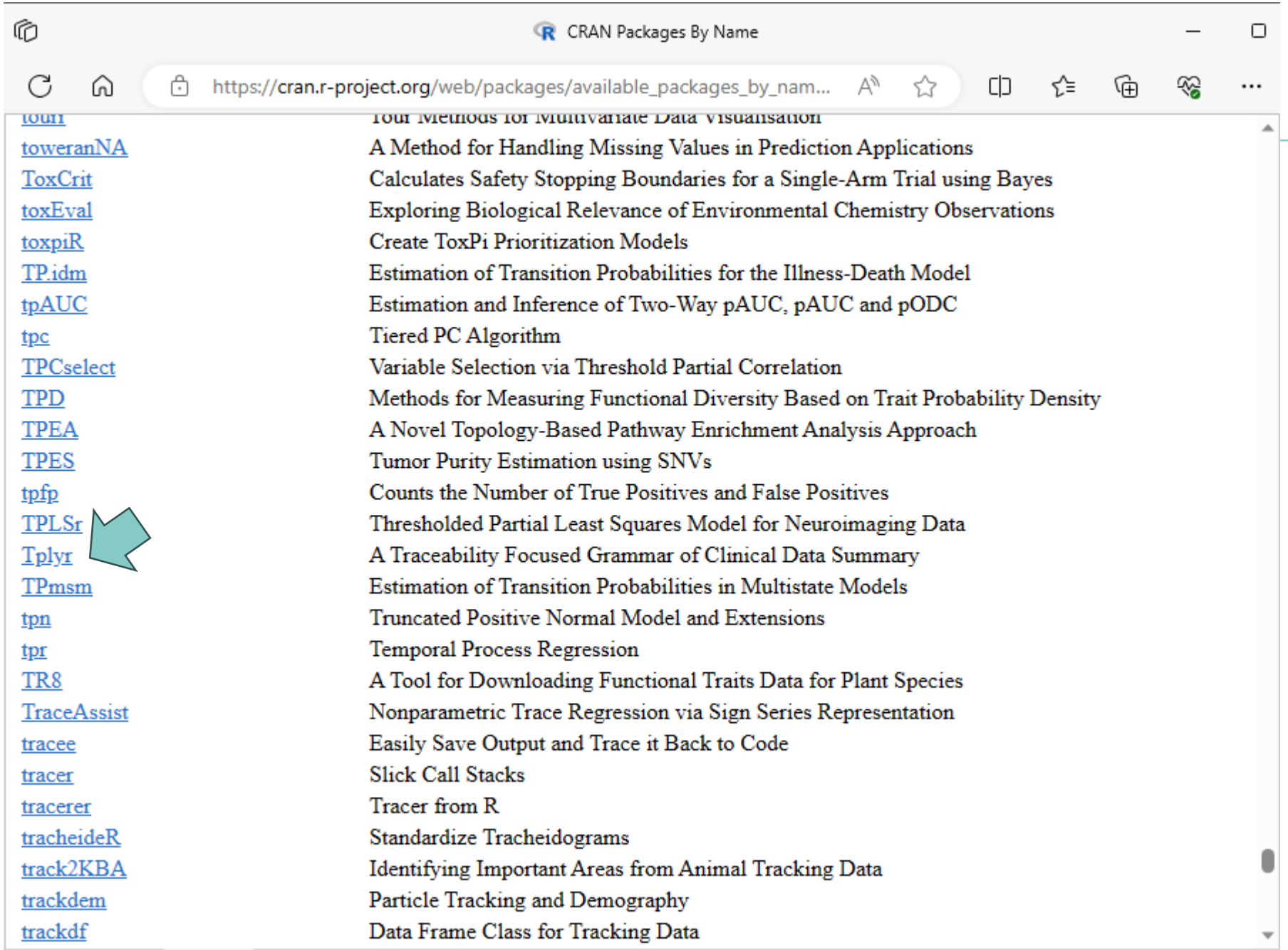
The results are summarized in the check summary (some timings are also available).

## Writing Your Own Packages

The manual Writing R Extensions (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

## Repository Policies

atorus
Your Complete Data Solution




| | |
|---|---|
| tourr | Tour Methods for Multivariate Data Visualisation |
| toweranNA | A Method for Handling Missing Values in Prediction Applications |
| ToxCrit | Calculates Safety Stopping Boundaries for a Single-Arm Trial using Bayes |
| toxEval | Exploring Biological Relevance of Environmental Chemistry Observations |
| toxpiR | Create ToxPi Prioritization Models |
| TP.idm | Estimation of Transition Probabilities for the Illness-Death Model |
| tpAUC | Estimation and Inference of Two-Way pAUC, pAUC and pODC |
| tpc | Tiered PC Algorithm |
| TPCselect | Variable Selection via Threshold Partial Correlation |
| TPD | Methods for Measuring Functional Diversity Based on Trait Probability Density |
| TPEA | A Novel Topology-Based Pathway Enrichment Analysis Approach |
| TPES | Tumor Purity Estimation using SNVs |
| tpfp | Counts the Number of True Positives and False Positives |
| TPLSr | Thresholded Partial Least Squares Model for Neuroimaging Data |
| Tplyr | A Traceability Focused Grammar of Clinical Data Summary |
| TPmsm | Estimation of Transition Probabilities in Multistate Models |
| tpn | Truncated Positive Normal Model and Extensions |
| tpr | Temporal Process Regression |
| TR8 | A Tool for Downloading Functional Traits Data for Plant Species |
| TraceAssist | Nonparametric Trace Regression via Sign Series Representation |
| tracee | Easily Save Output and Trace it Back to Code |
| tracer | Slick Call Stacks |
| tracerer | Tracer from R |
| tracheideR | Standardize Tracheidograms |
| track2KBA | Identifying Important Areas from Animal Tracking Data |
| trackdem | Particle Tracking and Demography |
| trackdf | Data Frame Class for Tracking Data |

atorus
Your Complete Data Solution

CRAN - Package Tplyr

https://cran.r-project.org/web/packages/Tplyr/index.html

# Tplyr: A Traceability Focused Grammar of Clinical Data Summary

A traceability focused tool created to simplify the data manipulation necessary to create clinical summaries.

| | |
|---|---|
| Version: | 1.1.0 |
| Depends: | R (≥ 3.5.0) |
| Imports: | rlang (≥ 0.4.6), assertthat (≥ 0.2.1), magrittr (≥ 1.5), dplyr (≥ 1.0.0), purrr (≥ 0.3.3), stringr (≥ 1.4.0), tidyr (≥ 1.0.2), tidyselect (≥ 1.1.0), tibble (≥ 3.0.1), lifecycle, forcats (≥ 0.4.0) |
| Suggests: | testthat (≥ 2.1.0), haven (≥ 2.2.0), knitr, rmarkdown, huxtable, tidyverse, readr, kableExtra, pharmaRTF, withr |
| Published: | 2023-01-10 |
| Author: | Eli Miller [aut], Mike Stackhouse [aut, cre], Ashley Tarasiewicz [aut], Nathan Kosiba [ctb], Sadchla Mascary [ctb], Atorus Research LLC [cph] |
| Maintainer: | Mike Stackhouse <mike.stackhouse at atorusresearch.com> |
| BugReports: | https://github.com/atorus-research/Tplyr/issues |
| License: | MIT + file LICENSE |
| URL: | https://github.com/atorus-research/Tplyr |
| NeedsCompilation: | no |
| Materials: | README NEWS |
| CRAN checks: | Tplyr results |

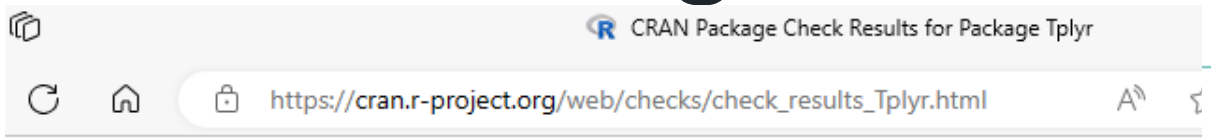
Documentation:

| | |
|---|---|
| Reference manual: | Tplyr.pdf |
| Vignettes: | Tplyr<br>count<br>custom-metadata<br>denom<br>desc_layers<br>desc_layer_formatting<br>general_string_formatting<br>layer_templates<br>metadata<br>tplyr_options<br>post_processing<br>shift |

atorus

Your Complete Data Solution

# CRAN Package Checks




**CRAN Package Check Results for Package Tplyr**

Last updated on 2023-08-29 13:53:36 CEST.

| Flavor | Version | $T_{install}$ | $T_{check}$ | $T_{total}$ | Status | Flags |
|---|---|---|---|---|---|---|
| r-devel-linux-x86_64-debian-clang | 1.1.0 | 15.43 | 285.30 | 300.73 | OK | |
| r-devel-linux-x86_64-debian-gcc | 1.1.0 | 10.14 | 210.40 | 220.54 | OK | |
| r-devel-linux-x86_64-fedora-clang | 1.1.0 | | | 390.44 | OK | |
| r-devel-linux-x86_64-fedora-gcc | 1.1.0 | | | 353.89 | OK | |
| r-devel-windows-x86_64 | 1.1.0 | 12.00 | 261.00 | 273.00 | OK | |
| r-patched-linux-x86_64 | 1.1.0 | 8.59 | 270.46 | 279.05 | OK | |
| r-release-linux-x86_64 | 1.1.0 | 8.00 | 272.89 | 280.89 | OK | |
| r-release-macos-arm64 | 1.1.0 | | | 122.00 | OK | |
| r-release-macos-x86_64 | 1.1.0 | | | 202.00 | OK | |
| r-release-windows-x86_64 | 1.1.0 | 17.00 | 285.00 | 302.00 | OK | |
| r-oldrel-macos-arm64 | 1.1.0 | | | 100.00 | OK | |
| r-oldrel-macos-x86_64 | 1.1.0 | | | 140.00 | OK | |
| r-oldrel-windows-x86_64 | 1.1.0 | 15.00 | 284.00 | 299.00 | OK | |

atorus
Your Complete Data Solution

# Validating Packages from CRAN

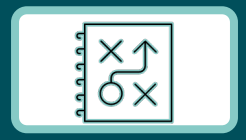
atorus

Your Complete Data Solution

# Validation Process

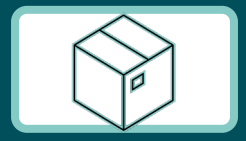
Discovery

Risk Assessment



Package Validation Plan

Testing and Validation



Package Release and Deployment

atorus

# Package Definitions

Data Processing: packages relate to cleaning, moving, and organizing data, they do not change or alter the data, rather reshape, and reorganize it.

Interactive: are any packages related to Shiny applications and interactive graphics

PK/PD: are any packages related to pharmacokinetics or pharmacodynamics

Statistics/ML: any packages that deal with frequentist or Bayesian statistics as well as any package related to supervised, unsupervised, or reinforcement machine learning methods.

Utility: involve any package that acts to help or facilitate the use of another package or data connection, or packages that produce a specific data format or output.

Visualization: are those that are related to the graphical display of data. Note that any visualization package that includes any statistical methods or is related to a PKPD is classified in that category

- R is installed from CRAN
- Includes verification and 15 ‘base packages’
- R was released in August 1993
- Derived from S which started in 1976 from Bell Labs

- Packages on CRAN are updated asynchronous to other packages
- Packages can have dependencies
- Vignettes or user instructions
- # of downloads
- Age of package

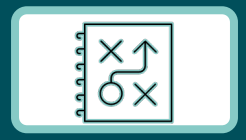
posit Package Manager

atorus
Your Complete Data Solution

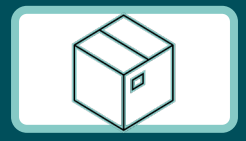
# Validation Process


Discovery


Risk Assessment


Package Validation Plan


Testing and Validation


Package Release and Deployment

atorus

# Risk Based Assessment

- The R Validation Hub: Implementing a Risk-Based Approach to R | R/Pharma (rinpharma.com)
- About Us (pharmar.org)


- Downloads as proxy
- References for functions
- Vignettes
- Source of package

- https://www.r-project.org/doc/R-FDA.pdf
- https://cran.r-project.org/web/packages/survival/vignettes/validate.pdf
- https://solutions.posit.co/envs-pkgs/environments/validation/
- https://www.rstudio.com/assets/img/validation-tidy.pdf

atorus
Your Complete Data Solution

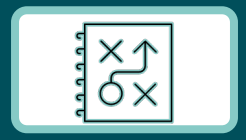
# Example Package Risk Assessment

| | |
|---|---|
| Package | ARTool |
| Version | 0.11.1 |
| Type | Statistics/ML |
| Maintainer | Matthew Kay <mjskay@northwestern.edu> |
| Description | The aligned rank transform for nonparametric factorial ANOVAs as described by Wobbrock, Findlater, Gergle, and Higgins (2011) <doi:10.1145/1978942.1978963>. Also supports aligned rank transform contrasts as described by Elkin, Kay, Higgins, and Wobbrock (2021) <doi:10.1145/3472749.3474784>. |
| Unit Tests Included | TRUE |
| Number.of.Dependencies | 6 |
| Dependencies | car, dplyr, emmeans, lme4, magrittr, plyr |
| Age | 8.7 |
| Vignettes | 2 |
| Downloads | 11802 |
| Download.Score | 0.07 |
| Download.Date | 07/31/2023 |
| Number.of.References | 2 |
| Risk.Level | Medium |

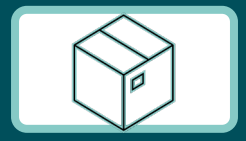
# Validation Process


Discovery


Risk Assessment


Package Validation Plan


Testing and Validation


Package Release and Deployment

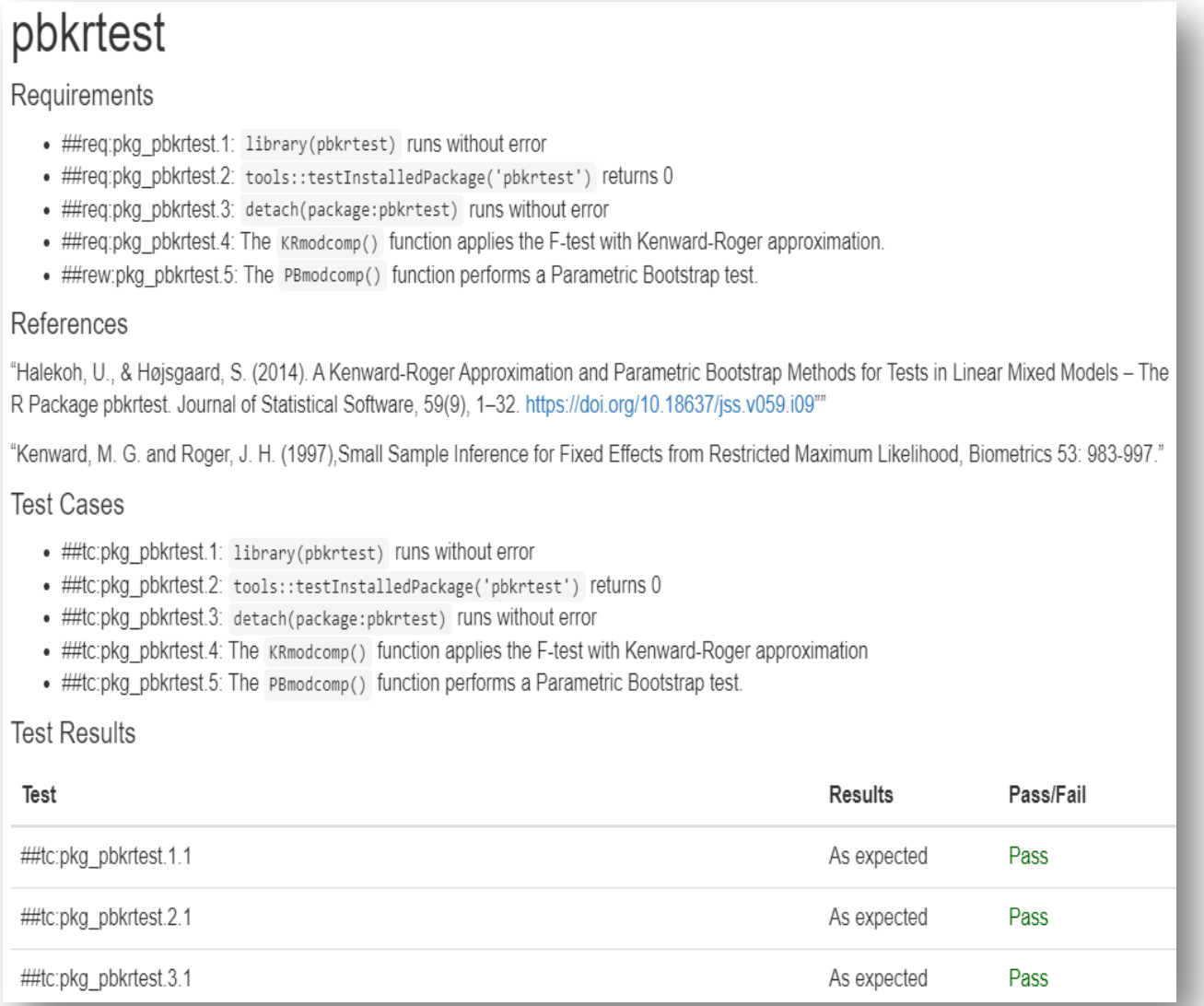
atorus

# Testing Scenarios

| Testing Mechanism | Explanation |
| --- | --- |
| Verified Output Comparison (Preferred) | Output comparison may also be done to compare programming output of a statistical test to known literature values. |
| Expectation Testing | A given test is expected to return some sort of message, warning, or error. The test will verify that in an expected situation, that message is returned properly. |
| Double Program | A programmer not involved with the development of the package code is used to produce the same result as the package code. This is done using a test scenario, where the package code is executed, and the programmer verifies the result of the test scenario value for value. |
| Snapshot Testing | For situations such as output files, an output is produced by the package. This output then undergoes human verification. After the output has been verified, it is then stored as a reference. Tests are then written to regenerate that output, and comparisons are run to make sure that new outputs files match the existing file exactly. |

# Per Package Validation Kit

- Package/function risk assessments
- Requirements established for packages which need additional testing
- Test cases written in English (i.e. “we are testing that the package does x”)
- Test code to verify those test cases, which is ultimately executed, and results are captured



## pbkrtest

### Requirements

- ##req:pkg_pbkrtest.1: `library(pbkrtest)` runs without error
- ##req:pkg_pbkrtest.2: `tools::testInstalledPackage('pbkrtest')` returns 0
- ##req:pkg_pbkrtest.3: `detach(package:pbkrtest)` runs without error
- ##req:pkg_pbkrtest.4: The `KRmodcomp()` function applies the F-test with Kenward-Roger approximation.
- ##rew:pkg_pbkrtest.5: The `PBmodcomp()` function performs a Parametric Bootstrap test.

### References

“Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest. Journal of Statistical Software, 59(9), 1–32. https://doi.org/10.18637/jss.v059.i09””

“Kenward, M. G. and Roger, J. H. (1997),Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood, Biometrics 53: 983-997.”

### Test Cases

- ##tc:pkg_pbkrtest.1: `library(pbkrtest)` runs without error
- ##tc:pkg_pbkrtest.2: `tools::testInstalledPackage('pbkrtest')` returns 0
- ##tc:pkg_pbkrtest.3: `detach(package:pbkrtest)` runs without error
- ##tc:pkg_pbkrtest.4: The `KRmodcomp()` function applies the F-test with Kenward-Roger approximation
- ##tc:pkg_pbkrtest.5: The `PBmodcomp()` function performs a Parametric Bootstrap test.

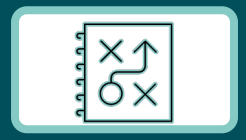
### Test Results

| Test | Results | Pass/Fail |
|---|---|---|
| ##tc:pkg_pbkrtest.1.1 | As expected | Pass |
| ##tc:pkg_pbkrtest.2.1 | As expected | Pass |
| ##tc:pkg_pbkrtest.3.1 | As expected | Pass |

# Validation Process

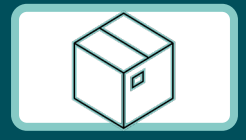
Discovery

Risk Assessment



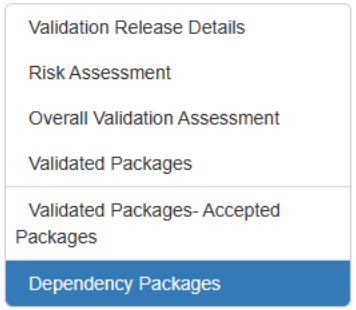
Package Validation Plan

Testing and Validation



Package Release and Deployment

# Testing and Validation Results




- Package Testing within Environment
- OS and SCE agnostic

## Overall Validation Assessment

| Package | N Tests | Tests Passed | Tests Failed |
|---|---|---|---|
| BH | 3 | 3 | 0 |
| BiasedUrn | 3 | 3 | 0 |
| BiocManager | 49 | 49 | 0 |
| BiocVersion | 3 | 3 | 0 |
| CBPS | 8 | 8 | 0 |
| DBI | 3 | 3 | 0 |
| DNAcopy | 3 | 3 | 0 |
| DT | 3 | 3 | 0 |
| DoseFinding | 28 | 28 | 0 |
| KernSmooth | 3 | 3 | 0 |

1–10 of 347 rows  Previous 1 2 3 4 5 ... 35 Next

## Dependency Packages

The following packages are dependencies for validated packages. These packages are tested to ensure they install correctly, all unit tests pass, and they detach properly.

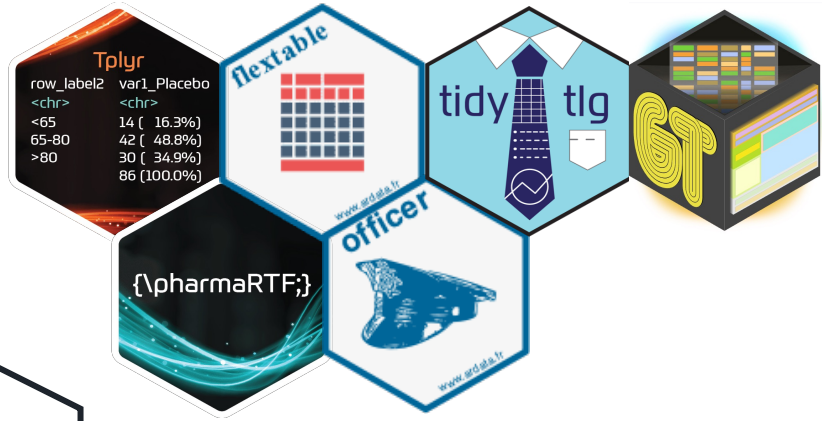
| Package | Version | Maintainer | Description | Tag |
|---|---|---|---|---|
| BH | 1.81.0.1 | dependency | | dependency |
| BiasedUrn | 2.0.8 | dependency | Statistical models of biased sampling in the form of univariate and multivariate noncentral hypergeometric distributions, including Wallenius' noncentral hypergeometric distribution and Fisher's | dependency |

atorus
Your Complete Data Solution

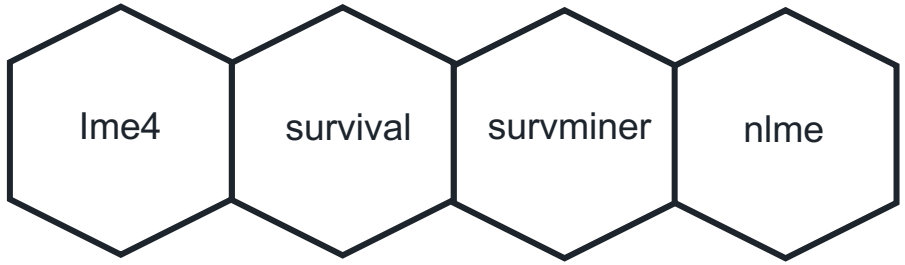
# Sampling of Packages Validated

## CDISC


## Tables

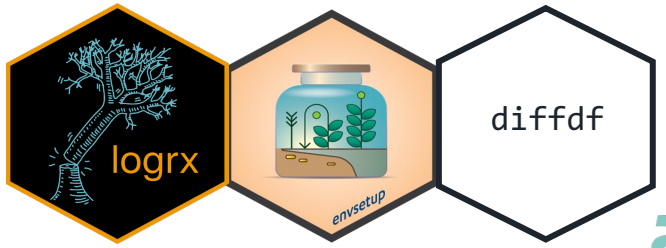



## Statistics




## Figures


## Utilities




atorus

Your Complete Data Solution

# Investments

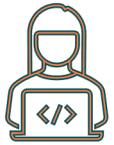
## 3 Years of Conversations
Validation foundation built on industry standards

## 3000+
Hours spent on development

## Collaboration
Internal and External
QA, IT, Statistics and Programming Teams

## 7+ Developers/Programmers
Implementing OpenVal for more than a year

## 8 Packages
Contribution and authorship on CRAN



## 500+
Test scripts written for 190 packages

## 40+
References cited

## Feedback Received
5+ industry Statisticians on package Inclusion and use cases for those packages

## Decades of deep-rooted clinical statistical programming expertise

posit™ phuse pharmaverse cdisc

# Thank You