Development and validation of open-source pharmacometrics software using an iterative software development life cycle methodology: the pmtables R package for submission-ready table generation

Kyle T. Baron, Devin Pastoor, Anna Nevison, Katherine Kay, Marc R. Gastonguay

Metrum Research Group, Tariffville, CT USA

Introduction

Scientists need standardized, programmatically generated tables summarizing outputs of modeling and simulation analyses. The objective of this work was to develop an R package to format typical pharmacometric output into LaTeX tables using an iterative open-source software development life cycle (SDLC) methodology.

Methods

pmtables is a freely available open-source R package, available on GitHub [1] under a general public license (GPL), for the summarization of data sets in publication-quality tables. The package was developed using a transparent, open SDLC process [2] that incorporated robust planning, iterative development, validation, and release.

In the planning of pmtables, the development team investigated the need for standardized tabular outputs in pharmacometric analyses. This exploratory phase included reviewing reports to identify standard table types and common structural features across tables that typically appear in reports. From this process, user stories were defined to represent desired features, and the user stories were then prioritized in a product roadmap.

Software development followed an iterative process of coding, testing, and feedback. Coding was performed incrementally, using the specifications and user stories prioritized in the planning process. Tests were written to closely reflect the real-world scenarios of the user stories and consisted of a series of unit tests that captured the expected behavior across pieces of code. Unit tests were executed locally by the developer as well as on a continuous integration server each time commits were made to the git repository. Feedback from the testing informed further coding, and this stage was considered complete after successful review of code, test results, and acceptance criteria. The iterative SDLC emphasizes multiple iterative cycles then feedback via user evaluation promotes an open source tool for optimal user experience and transparent engagement with community (**Figure 1**).
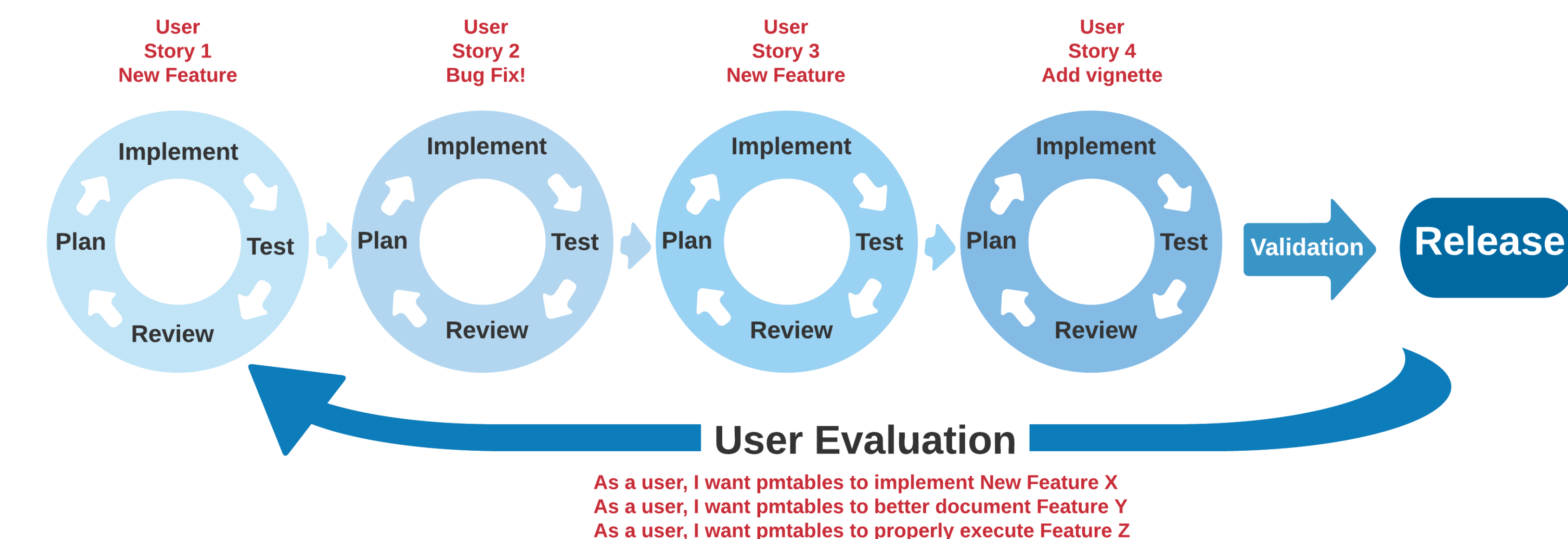

Figure 1: *Iterative SDLC*

Results

pmtables was released August, 2020 (v0.1.0), and six maintenance and development releases have been issued over the life of the package, including 133 issues submitted by pharmacometrics community end-users and closed by the development team. For the most recent release (v0.3.3), 606 expectations were evaluated by 206 unit tests (or user stories). The code coverage was 92%.

pmtables currently provides three data summary tables: continuous data summary, categorical data summary, and a data inventory table (summarizing numbers of individuals and observations in a data set). All tables can be rendered in a tabular environment or a longtable environment (the table can span multiple pages). Users have the option to (1) add footnotes (2) denote groups of rows (or panels) in the table (3) denote groups of columns with column "spanners" (4) add notes below the table and (5) add different types of styling to the table. Data set metadata (e.g., units, user-friendly names) can be extracted from a **yspec** data specification object [3] for direct use when building tables (**Table 2**). Tables can be annotated with the source R code file name and the output file name and automatically saved to .tex file in ascii format. Table preview tools are also provided to help users quickly visualize their tables in a test TeX document mimicking the report environment where the table will eventually be rendered. To help users learn the table syntax, an annotated example document was created and hosted on github [4].

```
tab <- pt_data_inventory(
    data,
    stacked = TRUE,
    by = c(Study = "STUDYf"),
    panel = c("Endpoint:" = "SEQf")
)
```

Study	Number SUBJ	MISS	OBS	BQL	Percent OBS	BQL
Endpoint: DEMO PK						
12-DEMO-001	30	8	427	15	13.4	0.5
12-DEMO-002	51	10	1152	38	36.3	1.2
11-DEMO-005	40	10	920	30	29.0	0.9
13-DEMO-001	40	7	582	11	18.3	0.3
Group Total	160	35	3081	94	97.0	3.0
Endpoint: ESTRDIOL						
11-DEMO-005	40	0	40	0	50.6	0.0
13-DEMO-001	40	1	39	0	49.4	0.0
Group Total	80	1	79	0	100.0	0.0
Endpoint: BMD						
11-DEMO-005	40	9	111	0	49.1	0.0
13-DEMO-001	40	5	115	0	50.9	0.0
Group Total	80	14	226	0	100.0	0.0

SUBJ: subjects
BQL: below quantitation limit
MISS: missing observations (not BQL)
OBS: observations

Table 1: *Data Disposition Summary*

```
short <- ys_get_short_unit(spec, parens = TRUE, title_case = TRUE, short_max = 8)

tab <- pt_cont_wide(
    data = data,
    cols = "WT,SCR,AGE,ALB,HT",
    by = c(Study = "STUDYf"),
    panel = as.panel("FORMf", prefix = "Formulation: "),
    table = short
)
```

Study	Weight (kg)	SCR (mg/dL)	Age (years)	Albumin (g/dL)	Height (cm)
Formulation: tablet					
12-DEMO-001	71.0 (14.2) [24]	1.01 (0.157) [25]	32.6 (9.23) [25]	4.22 (0.459) [24]	179 (19.7) [25]
12-DEMO-002	72.2 (11.8) [41]	0.966 (0.166) [42]	34.0 (7.93) [42]	4.49 (0.495) [42]	182 (15.9) [42]
11-DEMO-005	68.8 (15.2) [29]	2.48 (1.47) [30]	33.2 (8.73) [30]	4.37 (0.568) [29]	173 (19.7) [30]
13-DEMO-001	69.4 (11.0) [33]	0.967 (0.163) [33]	33.7 (9.67) [33]	3.53 (1.14) [31]	178 (16.5) [33]
Formulation: capsule					
12-DEMO-001	72.9 (17.3) [3]	1.12 (0.0700) [3]	32.2 (12.0) [3]	4.49 (0.593) [3]	184 (23.0) [3]
12-DEMO-002	70.9 (10.3) [6]	1.03 (0.146) [6]	37.7 (7.59) [6]	4.38 (0.354) [6]	181 (15.4) [6]
11-DEMO-005	73.9 (11.1) [3]	3.06 (2.19) [3]	31.8 (4.99) [3]	4.65 (0.240) [3]	181 (16.4) [3]
13-DEMO-001	58.4 (4.04) [3]	0.973 (0.195) [3]	36.5 (6.69) [3]	3.09 (1.50) [3]	167 (8.88) [3]
Formulation: troche					
12-DEMO-001	85.3 (12.4) [2]	1.20 (0.0707) [2]	25.1 (3.28) [2]	4.74 (0.283) [2]	194 (0.163) [2]
12-DEMO-002	79.7 (8.61) [2]	0.910 (0.0283) [2]	48.0 (1.79) [2]	4.49 (0.0354) [2]	182 (10.9) [2]
11-DEMO-005	66.8 (13.9) [7]	2.45 (1.05) [7]	31.4 (9.34) [7]	4.49 (0.509) [7]	177 (19.8) [7]
13-DEMO-001	77.4 (15.9) [4]	0.795 (0.0777) [4]	37.3 (12.9) [4]	4.32 (0.994) [4]	193 (22.4) [4]
All data	70.7 (12.8) [157]	1.36 (0.986) [160]	33.7 (8.83) [160]	4.20 (0.793) [156]	179 (17.7) [160]

Summary is mean (sd) [count]

Table 2: *Continuous Data Summary*

```
cols <- c(Sex = "SEXf", Race = "ASIANf", Formulation = "FORMf", `Renal function` = "RFf")

tab <- pt_cat_long(
    data,
    cols = cols,
    span = c("Study Number" = "STUDYf")
)
```

	Study Number 12-DEMO-001 n = 30	12-DEMO-002 n = 50	11-DEMO-005 n = 40	13-DEMO-001 n = 40	Summary n = 160
Sex					
male	10 (33.3)	18 (36.0)	29 (72.5)	23 (57.5)	80 (50.0)
female	20 (66.7)	32 (64.0)	11 (27.5)	17 (42.5)	80 (50.0)
Race					
Asian	17 (56.7)	18 (36.0)	16 (40.0)	15 (37.5)	66 (41.2)
non-Asian	13 (43.3)	32 (64.0)	24 (60.0)	25 (62.5)	94 (58.8)
Formulation					
tablet	25 (83.3)	42 (84.0)	30 (75.0)	33 (82.5)	130 (81.2)
capsule	3 (10.0)	6 (12.0)	3 (7.5)	3 (7.5)	15 (9.4)
troche	2 (6.7)	2 (4.0)	7 (17.5)	4 (10.0)	15 (9.4)
Renal function					
normal	30 (100.0)	50 (100.0)	10 (25.0)	40 (100.0)	130 (81.2)
mild	0 (0.0)	0 (0.0)	10 (25.0)	0 (0.0)	10 (6.2)
moderate	0 (0.0)	0 (0.0)	10 (25.0)	0 (0.0)	10 (6.2)
severe	0 (0.0)	0 (0.0)	10 (25.0)	0 (0.0)	10 (6.2)

Summary is count (percent)
n: number of records summarized

Table 3: *Categorical Data Summary*

Conclusion

The development and validation of the pmtables package illustrates the feasibility and utility of a modern SDLC for open source pharmacometrics software.

References

[1] https://github.com/metrumresearchgroup/pmtables
[2] https://github.com/metrumresearchgroup/open-sdlc
[3] https://github.com/metrumresearchgroup/yspec
[4] https://github.com/metrumresearchgroup/pmt-book