

# オープンソースソフトウェアの活用と留意事項

日本製薬工業協会 医薬品評価委員会 データサイエンス部会

2022年度 タスクフォース 5

2023年4月

## 目次

略語一覧表.....	3
はじめに .....	4
1. オープンソースソフトウェア .....	4
1.1 オープンソースソフトウェアのライセンス .....	6
1.2 統計解析用ソフトウェア「R」 .....	7
1.3 製薬業界における R に係る活動.....	8
2. OSS を承認申請で利用する上での課題.....	11
2.1 運用業務の管理.....	11
2.2 R のバージョン管理.....	12
2.3 R パッケージのバージョン管理.....	12
2.3.1 R パッケージのバージョンを固定して運用する場合 .....	13
2.3.2 R パッケージをマルチバージョンで運用する場合 .....	13
2.4 R パッケージマネージメント .....	13
2.5 管理体制.....	14
2.5.1 R パッケージの管理体制 .....	14
3. R の教育研修 .....	16
3.1 R トレーニング事例.....	16
3.2 R トレーニングのカリキュラム.....	17
3.3 R 学習参考資料 .....	18
おわりに .....	19

## 略語一覧表

略語	略していない表現 (英語)	略していない表現 (日本語)
OSS	Open Source Software	オープンソースソフトウェア
OSI	Open Source Initiative	オープンソースを促進することを目的とする組織
OSD	The Open Source Definition	OSI が策定したオープンソースとされるものが満たすべき条件を示した文書
CRAN	Comprehensive R Archive Network	包括的 R アーカイブネットワーク

## はじめに

これまで臨床開発におけるデータマネジメントやデータ解析の業務には、ソフトウェアベンダーが開発した各種の「プロプライエタリソフトウェア」が多く使用されてきた。プロプライエタリソフトウェアとは、ソフトウェアの開発者がソースコードを独占的に保持し、ソフトウェアのユーザにはソースコードが通常開示されないソフトウェアのことである。透明性が求められる臨床開発の業務でプロプライエタリソフトウェアが使用されてきた背景には、ソフトウェア開発企業がユーザからの機能追加や仕様の説明への要求に対応し、それによってユーザの一部である製薬会社も自らの業務を問題なく遂行し、規制当局にも業務の内容を説明することができていたからだと考えられる。

しかし、近年、臨床開発の分野で「オープンソースソフトウェア」が使用される機会が増えてきた。オープンソースソフトウェアとは、プロプライエタリソフトウェアとは反対に、ソフトウェアのユーザにソースコードが開示され、ユーザもソースコードを自らの目的に合わせて改変できるソフトウェアのことである（詳細は「1. オープンソースソフトウェア」を参照）。このようなソフトウェアが臨床開発で使用されるようになってきた背景には、ソフトウェアが世界中の開発者によって改良されるため新しい技術や統計手法が比較的速やかに実現できる、プロプライエタリソフトウェアに比べてコストが抑えられる、また、製薬会社に人材を輩出している大学等の教育機関でもオープンソースソフトウェアを使うことが増えているため、オープンソースソフトウェアを使用できる環境を社内に整えて新たに獲得した人材にも活躍してほしい、といった期待があると考えられる。

本報告書では、オープンソースソフトウェアの定義を最初に紹介し、その後、製薬業界での取り組みが近年特に活発になっているオープンソースソフトウェアと、その長所や使用上の注意点を紹介する。さらに、オープンソースソフトウェアを承認申請業務に使用することを想定した場合に必要なと考えられる管理体制や、教育研修の内容を、Rを事例として紹介する。本報告書がオープンソースソフトウェア、特にRを社内で使用する際の検討の一助になれば幸いである。

## 1. オープンソースソフトウェア

オープンソースソフトウェア（OSS）の定義は幾つかあるものの、総務省は平成18年版情報通信白書<sup>1)</sup>において、「インターネットを通じてソースコードが公開され、誰でも自由に使用・改変・再配布できるソフトウェア」、「厳密には Open Source Initiative（OSI）が定めた「The Open Source Definition（OSD）」と呼ばれる定義を満たすソフトウェア」のことであると記載している。

OSIはオープンソースソフトウェアの促進を目的とする組織であり、OSDとしてOSSの定義を以下の通りに記載している<sup>2)</sup>（日本語記載は Open Source Group Japan の和訳<sup>3)</sup>を記載した）。

- Free Redistribution (再頒布の自由)
- Source Code (ソースコード)
- Derived Works (派生ソフトウェア)
- Integrity of The Author's Source Code (作者のソースコードの完全性)
- No Discrimination Against Persons or Groups (個人やグループに対する差別の禁止)
- No Discrimination Against Fields of Endeavor (利用する分野に対する差別の禁止)
- Distribution of License (ライセンスの分配)
- License Must Not Be Specific to a Product (特定製品でのみ有効なライセンスの禁止)
- License Must Not Restrict Other Software (他のソフトウェアを制限するライセンスの禁止)
- License Must Be Technology-Neutral (ライセンスは技術中立的でなければならない)

OSS は上記の条件に記載された特徴があるため、「ソフトウェア価格の大幅な低下の可能性」<sup>1)</sup>「誰でも自由に利用可能で、誰でもソースコードの解析・研究が可能」<sup>4)</sup>といった期待があり、既に多くの分野で以下のようなソフトウェアが使用されている。

表 1 オープンソースソフトウェアの例

分類	ソフトウェア名
オペレーティングシステム	Linux, FreeBSD, Darwin
インターネットサーバ	Apache (WWW サーバ), BIND (DNS サーバ), Sendmail (メールサーバ), Zope (アプリケーションサーバ)
データベース	MySQL, PostgreSQL
デスクトップ統合環境	GNOME, KDE
デスクトップソフトウェア	Mozilla (ブラウザ), StarOffice/OpenOffice.org (オフィススイート), GIMP (グラフィックエディタ)
スクリプト言語	Perl, Python, R, Ruby, Tcl/Tk

ソフトウェア情報センター「オープンソースソフトウェアの現状と今後の課題について」より作成

#### 参考文献

- 1) 総務省 平成 18 年版情報通信白書  
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h18/html/i1452000.html>
- 2) Open Source Initiative “The Open Source Definition Version 1.9,”.  
<https://opensource.org/docs/osd>
- 3) Open Source Group Japan “オープンソースの定義 (v1.9)”  
<https://opensource.jp/osd/osd19plain/>

#### 4) 情報処理推進機構 “OSS の概要に関する知識”

<https://warp.ndl.go.jp/info:ndljp/pid/12446699/www.ipa.go.jp/files/000056269.pdf>

##### 1.1 オープンソースソフトウェアのライセンス

OSS には幾つかの種類のライセンスがあり、以下は代表的なものである<sup>1)</sup>。

###### ● GNU GPL (General Public License)

- リチャード・ストールマン氏によって考案されたライセンス契約である。
- 自由に利用、再配布、改変や派生ソフトウェアの作成・配布が可能である。
- 派生ソフトウェア（改変したソフトウェアあるいは一体化して用いるソフトウェア）についても GNU GPL ライセンスに基づきソースコードの公開、利用、再配布、改変等が定められている。

###### ● BSDL (Berkeley Software Distribution License)

- 米国カリフォルニア州立大学バークレー校により開発されたライセンス契約である。
- 利用や再配布にあたり、著作権表示ならびに、無保証であり、自己責任で使用する事への同意（免責条項）を許諾すればよく、制限が少ない。
- ソースコード非公開の商用ソフトウェアへの取り込みが可能である。

###### ● MPL (Mozilla Public License)

- ネットスケープ社が Netscape を OSS 化した NPL(Netscape Public License)をもととしている。
- GNU GPL と同じく、改変したソフトウェアを再配布する場合にはソースコードの公開が義務付けられている。
- 独立した自己開発ソフトウェアと MPL ソフトウェアを一体化し、ひとつのプログラムとする場合、GNU GPL では自己開発部分もソースコードの公開義務があるが、MPL では非公開とするなど別のライセンスを採用できる。

上記の代表的なライセンスを含め、OSS のライセンスによっては改変したソフトウェアを再配布する際にソースコードを公開するといった制限はあるものの、1.1 の OSD にある通り、単に利用するだけであれば制限はない。そのため、製薬会社がソフトウェアを再配布することなく、自社の業務のために OSS を利用することは問題ないと考えられる。しかし、プログラム医療機器のプログラムに OSS が含まれる場合等、製薬会社が OSS を含んだソフトウェアを開発して配布する際には、上記のライセンスに従ってソースコードを公開すると

いった対応が必要になる場合があると考えられる。

#### 参考文献

- 1) 情報処理推進機構 “法務分野に関する知識（基本レベル）”

<https://warp.ndl.go.jp/info:ndljp/pid/12446699/www.ipa.go.jp/files/000056274.pdf>

### 1.2 統計解析用ソフトウェア「R」

GNU GPL のライセンスに従うソフトウェアの一つが R である。R とは、統計解析用ソフトウェア「S」で使用する「S 言語」が動作することを目的に作成されたオープンソースの統計解析用フリーソフトウェアである<sup>1)</sup>。

R は UNIX 系、Windows、MacOS の 3 種類の OS に対応しており<sup>2)</sup>、ソフトウェアをインストールしたのと同時に利用できる統計解析機能に加え、「パッケージ」と呼ばれる統計解析手法ごとのプログラムのまとまりを、インターネットを介してダウンロードすることで統計解析機能を拡張できる<sup>1)</sup>。R のソフトウェアは R Foundation によって管理されており、「CRAN (Comprehensive R Archive Network)」と呼ばれるウェブサイトで公開されている<sup>3)</sup>。パッケージは全世界の開発者によって作成されており、通常は CRAN で公開される。

R のパッケージには、伝統的な統計解析手法に対応するものだけでなく、ベイズ統計、因果推論、機械学習といった新しい手法に対応するもの、レポート作成機能や高度なグラフ作成を実現するもの、Microsoft Excel や Word、SAS や SPSS といった他のソフトウェアのファイルの入出力を行うものまで、多くの種類がある。現在、CRAN に公開されているパッケージだけでも約 20,000 種類がある。

他のソフトウェアと同様、R も機能追加やバグの修正のために不定期にソフトウェアがバージョンアップされる。R パッケージも同様にバージョンアップが行われるが、R のバージョンアップとタイミングを合わせているわけではない。

#### 参考文献

- 1) R Foundation “What is R?”

<https://www.r-project.org/about.html>

- 2) R Foundation “What machines does R run on?”

[https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-machines-does-R-run-on\\_003f](https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-machines-does-R-run-on_003f)

- 3) R Foundation “CRAN Mirrors”

<https://cran.r-project.org/mirrors.html>

### 1.3 製薬業界における R に係る活動

R には、無償であることに加え、世界中のユーザがパッケージを作成・更新するため最先端の解析手法が直ぐに使える、また、近年開発された可読性が高いプログラムの記載を実現できるパッケージ「tidyverse」がある、といった長所がある。このような R の利便性から、製薬業界でも臨床試験データの分析や、そのデータマネジメントに係る業務に活用するための取り組みが始まっている。

近年では、承認申請を目的とした業務への活用の検討が特に活発になってきた。その代表的な団体を以下で紹介する。

#### 1. R submission working group

R consortium<sup>1)</sup>に所属している製薬業界メンバーで構成されており、承認申請時に R ベースの承認申請時の電子データを規制当局へ提出を目標に活動している。

現在、FDA と協力してパイロットを進めており、2021 年 11 月に実施された 1 回目のパイロットサブミッションでは、R で作成された解析帳票とそのプログラム、解析帳票の再現性確保のための項が追加された ADRG (Analysis Data Reviewer's Guide) を含む申請電子データパッケージを作成し、パイロットサブミッションを実施。提出資料はすべて GitHub 上で公開<sup>2)</sup>されており、FDA での確認結果も本サイトで確認できるようになっている。また、R Consortium の webinar にて実施された、FDA のレビューアによる本パイロットのフィードバックも YouTube にて公開<sup>3)</sup>されている。2022 年 9 月から開始されたパイロットサブミッションでは、1 回目の申請データパッケージに加え、インタラクティブなデータレビューを目的とした R Shiny を含む申請パッケージのパイロットサブミッションを実施し、FDA により受領されたことが報告されている<sup>4)</sup>。

#### 参考文献

##### 1) R consortium

全世界の R ソフトウェアユーザや、開発者（メンテナンスも含む）のコミュニティをサポートするために組織されたグループで、R プロジェクトの管理機関である R Foundation のサポートをしている。

<https://www.r-consortium.org/>

##### 2) 1 回目パイロットサブミッション提出資料

<https://github.com/RConsortium/submissions-pilot1-to-fda>

##### 3) R Consortium webinar 「Using R in Regulatory Review

<https://www.r-consortium.org/using-r-in-regulatory-review>

##### 4) 2 回目パイロットサブミッションに関する情報 (UPDATE: Successful R-Based Package Submission with Shiny Component to FDA)

<https://www.r-consortium.org/blog/2022/12/07/update-successful-r-based-package-submission-with-shiny-component-to-fda>

## 2. Pharmaverse

製薬企業からの参加者で構成された団体で、CRF から臨床試験データの分析レポートの提出までをサポートする R パッケージの開発とメンテナンスを実施<sup>1)</sup>。代表的な成果物 (R パッケージ) を以下に紹介する。

- sdtmcheck  
SDTM のデータチェック用 R パッケージ。詳細な検査は以下の URL を参照。  
<https://pharmaverse.github.io/sdtmchecks/reference/index.html#data-checks>
- admiral  
ADaM 作成用 R パッケージ。ADaM を作成するために必要な機能を関数として提供している。  
<https://pharmaverse.github.io/admiral/cran-release/index.html>

admiral は、がん領域 (admiralonco)、眼科領域 (admiralophtha)、ワクチン領域 (admiralvaccine) の有効性評価用の ADaM を開発するために必要な機能を提供している。

admiralonco: <https://pharmaverse.github.io/admiralonco/main/index.html>

admiralophtha: <https://pharmaverse.github.io/admiralophtha/main/>

admiralvaccine (開発中) : <https://github.com/pharmaverse/admiralvaccine>

- rtable  
データから帳票の体裁に整形して出力する R パッケージ。  
<https://insightengineering.github.io/rtables/main/>
- tern  
要約統計量や統計処理 (信頼区間、回帰分析、分散分析、生存時間解析等) を実行する R パッケージ。  
<https://insightengineering.github.io/tern/latest-tag/>
- teal  
Rshiny のアプリケーション開発をサポートする R パッケージ。  
<https://insightengineering.github.io/teal/latest-tag/>

- logrx  
R プログラムの実行ログファイルを作成する R パッケージ。  
<https://pharmaverse.github.io/logrx/index.html>
- xportr  
R data frame 形式の ADaM や SDTM を、xpt ファイルに変換する R パッケージ。xpt ファイルへの変換時に SAS ver.5 移送形式の要件※に対応しているかのチェックも行う。  
<https://atorus-research.github.io/xportr/>

※ SAS ver.5 移送形式の要件

[https://documentation.sas.com/doc/en/pgmsascdc/9.4\\_3.5/movefile/n1xbwdre0giahfn11c99yjkpi2yb.htm](https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/movefile/n1xbwdre0giahfn11c99yjkpi2yb.htm)

#### 参考文献

- 1) Pharmaverse  
<https://pharmaverse.org/charter/>
3. R validation Hub  
医薬品の規制環境における R の導入支援を目的とした、製薬企業からの参加者で構成された団体で、R を承認申請のために利用すべく、R パッケージを利用目的 (statistical / non-statistical)、メンテナンス状況、コミュニティの利用状況を元にリスクアセスメントする事で、その信頼性を評価し、パッケージの採用・不採用や、リスク軽減措置を講じる方法<sup>1)</sup>が検討されている。

#### 参考文献

- 1) R Validation hub  
<https://www.pharmar.org/>  
A RISK-BASED APPROACH FOR ASSESSING R PACKAGE ACCURACY  
WITHIN A VALIDATED INFRASTRUCTURE  
<https://www.pharmar.org/white-paper/>

## 2. OSS を承認申請で利用する上での課題

FDA や PMDA といった規制当局は、承認申請を目的とした臨床試験データの分析や、そのデータマネジメントに係る業務において、特定のソフトウェアを利用する事を要求していない。このため利用するソフトウェアの選択は申請者に委ねられており、OSS を承認申請目的として利用することも問題ないと考えられる。OSS はこれまで承認申請業務に広く用いられてきた SAS や SPSS と比較し、ユーザが自由に使用でき、導入に対する敷居が低く機能拡張の速度も速い。また高度なプログラミング技術を必要としない OSS もあり、応用の幅も広いことが考えられる。しかし、自由度の高さ故に、申請者でソフトウェアの品質（言い換えるならば、解析結果の信頼性）を確保するための検証作業を実施する必要がある、その手順や結果の文書化が求められる（ICH E9 「5.8 Integrity of Data and Computer Software Validity」<sup>1)</sup>）。OSS を利用する場合であっても同様の対応が必要である。

一般的に、これらの信頼性の確保には、以下の項目を保証するための手順や結果が文書化されている必要がある。

- ・ Accuracy (信頼性)
- ・ Reproducibility (再現性)
- ・ Traceability (追跡可能性)

Traceability に関しては、データやプログラムが保管されている解析システムや、その決められた利用手順に従う事により担保され得るものだが、Accuracy に関しては、たとえ適切な手順で解析結果自体の品質保証が実施されていても、解析に使用した OSS の信頼性が低ければ、その解析結果の信頼性自体が疑われることになる。また、Reproducibility においては、解析結果と共に、それに利用された OSS のバージョンが管理され、解析結果を再現可能なバージョンを提示する必要がある。以降の章では、R の事例を元に Accuracy や Reproducibility を担保すべく、OSS の管理やその運用方法について紹介する。

### 参考文献

#### 1) ICH E9

<https://www.pmda.go.jp/files/000156905.pdf>

### 2.1 運用業務の管理

承認申請で R を利用する際は前述の各プロセスにおける SOP やそれに準じる作業手順書の整備が必要となる。

また、R の利用環境においてクラウドサーバーを利用するケースも増えており、その場合

はクラウドサーバーの管理体制を IT 部門にて準備する必要がある。

各解析業務の実施の管理においては、使用した R および R パッケージのバージョンの履歴の記録が必要となり、その管理体制が必要となる。

## 2.2 R のバージョン管理

R は不定期にバージョンアップされるため、Reproducibility の確保のためには、少なくとも CSR 用解析から承認申請取得までは一貫したバージョンを利用するのが望ましい。しかしながら、個々の CSR 分析から承認申請までの期間は、長いものでは 10 年以上もあり、その間にバグ修正版のリリースへの対応や、最新 R パッケージとのバージョンのミスマッチ（R パッケージに対して利用する R のバージョンが古すぎて参照できない）といった問題が発生する。このため、CSR 解析時点の R とその R パッケージのバージョンのセットと、解析時点から更新された R のバージョンと、それに対応した R パッケージのバージョンのセットを複数維持する必要があり、社内でのパッケージの管理が非常に煩雑になる問題が発生する。この問題を解消する方法として、R のバージョンに対する凍結レポジトリを運用する方法が Posit (旧 RStudio) の [website<sup>1\)</sup>](https://solutions.posit.co/envs-pkgs/environments/)にて紹介されている。本運用により、特定バージョンの R に対する、パッケージの管理がシンプルなものになり、R パッケージのバージョンアップにも対応が可能になる。

## 参考文献

### 1) Posit

Reproducible Environments

<https://solutions.posit.co/envs-pkgs/environments/>

## 2.3 R パッケージのバージョン管理

ほぼすべての R パッケージは他の R パッケージを参照しており、これらの参照関係はパッケージの依存関係と呼ばれる。CRAN の website には、各 R パッケージの依存関係についても記載されている。

複数の R パッケージをインストールする場合や、組織内の全員が確実に正しい R パッケージの種類とバージョンを取得するためには、R パッケージ間の複雑な依存関係を理解し、正しいバージョンを使用することが重要であり、バージョンと依存関係を追跡することで、R プログラムの実行結果の再現性を確保できる。しかしながら、個々の R パッケージのバージョンと、依存関係のある R パッケージのバージョンを一度に把握し管理するには、相応の管理体制が必要となる。

本項では、バージョン管理の方法として、R パッケージのバージョンを固定して運用する方法と、マルチバージョンで R パッケージを管理しつつも、プロジェクト内で一貫した R パッケージのバージョンを使って運用する方法と注意すべき事項について紹介する。

### 2.3.1 R パッケージのバージョンを固定して運用する場合

同一プロジェクト内において複数人でプログラミングを実施するためには効率的な方法となる。また、頻繁にパッケージのバージョンアップを実施しないため、バージョン管理は単純化される。ただし、頻繁にバージョンアップをしないため、機能面の不足や、未対応のバグに対応していないなどの不都合が発生する。さらには、固定バージョンからバージョンアップが必要になった場合に、パッケージの依存関係の修正等により、余計なメンテナンスコストが発生する可能性もあるため注意したい。

### 2.3.2 R パッケージをマルチバージョンで運用する場合

マルチバージョンのパッケージが存在しているケースでは、“renv”<sup>2)</sup>という機能を使用することができる。“renv”とは Posit (旧 RStudio)が開発している R パッケージ管理のためのパッケージであり、バージョンの同じ R パッケージを使用することができる。“renv”には R パッケージのバージョンとその R パッケージを取得できる外部ソースを記録することができる。これらの情報により、既存プロジェクトの R パッケージを別プロジェクトにて再現することができる。ただし、必ずしも“renv”の機能のみで再現性を確保できる訳ではない。実行環境であるオペレーティングシステム内の他のコンポーネント (Windows システムパッケージ, R のコンパイラ等) に依存しており、また、非常に稀にプロジェクトに依存関係のある一部の R パッケージが利用できなくなり CRAN から将来的に削除される可能性もある。

#### 参考文献

1) CRAN

<https://cran.ism.ac.jp/web/packages/index.html>

2) renv

<https://rstudio.github.io/renv/articles/renv.html>

## 2.4 R パッケージマネージメント

R パッケージ自体は誰でも自由に配布することが出来るため、その品質の担保は重要な課題となる。CRAN で公開されている R パッケージは、配布前にパッケージのテストは行われているが、全てのバグが解消されているとは限らない。また有名な repository の一つである GitHub にも多くの R パッケージが公開されているが、内容のレビューは行われていない。

独自開発されたパッケージのアップデートの有無を確認するためには、`update.packages()`や `old.packages()` コマンドを用いることで、最新のバージョンへのアップデートやアップデート可能な古いパッケージを確認することが出来る。

CRAN には、約 20,000 近くの R パッケージが公開されており、その一覧が CRAN のウェブページ上に公開されている(2022 年 12 月時点)。CRAN Task view では、特定のトピックに関連する CRAN 上のパッケージについてガイダンスが提供されており、ベイズ推定や生存分析、薬物動態などのトピックについてもまとめられている。但し、記載されている R パッケージが必ずしも適切である保証はない点に留意する必要がある。また、Posit (旧 RStudio) の website にも CRAN で公開されている R パッケージのうち、一般的で有用なものが解説付きでまとめられている。

#### 参考文献

1) CRAN

<https://cran.r-project.org/>

2) Posit

<https://posit.co/>

### 2.5 管理体制

承認申請において R を活用する際は、その解析結果の信頼性を保証する上で様々な管理体制が必要となる。本項では想定し得る管理体制について記載した。

#### 2.5.1 R パッケージの管理体制

R validation Hub<sup>1)</sup> では R パッケージの信頼度と利用目的に応じたリスク別の運用を提言している。本項ではこの提言に基づく運用事例について紹介する。

- ノバルティス社での事例<sup>2)</sup>

ノバルティス社では以下の 2 つの use case および R パッケージのリスク評価に基づいた管理体制をとっている。

use case1 : 標準の R パッケージとして医薬品開発の通常業務で使用され、プラットフォーム上でプレインストールされるもの

use case2 : 特定のプロジェクトのみで使用される R パッケージ

R パッケージのリスク評価については、R validation Hub の提言に基づいたチェックリストが用意されており、R、利用できる R パッケージを低リスク、中リスク、高リスクの 3 段階に分類している。基準に満たない R パッケージについては極めて高リスクと分類されたものについては原則利用不可としている。

use case1 の場合:

低リスクまたは中リスクと分類された R パッケージについては検証作業を不要とし、IT 部門が R 利用環境にプレインストールし、通常業務にて利用される。

高リスクと評価された R パッケージについては IT 部門により必要な検証プロセスを経て問題なければ利用できるとされている。また、この場合、QA 部門での承認が必要となる。

use case2 の場合:

プロジェクトチームにより R パッケージの検証、記録の管理が行われる。このプロセスにおいては自動化ツールが用意されており、これを用いて R パッケージのインストール、テストスクリプトの実行と検証、検証記録の保存等が可能となっている。低リスク、中リスクに分類された R パッケージはこの自動化ツールにより管理が行われ、高リスクに分類された場合は加えてエンドユーザーによる検証が必要となる。

● メルク社での事例<sup>3)4)</sup>

メルク社では、R パッケージのリスクを評価する基準を設けており、その基準を満たせば低リスクまたは中リスクに分類される。また、いずれの基準にも満たなかった場合はオープンと分類され、それぞれのリスク分類と利用目的に応じて管理されている。リスク分類と利用目的の関連性は以下の表の通りとなる。

利用用途	利用可能な R パッケージ分類
外部への提供(eCTD)	低リスク
外部への提供(eCTD 以外) 社内利用 (部署外)	中リスクまたは低リスク
探索的な用途/部署内での利用	オープン、中リスク、低リスク

R パッケージ管理では RStudio が推奨している Shared Baselines<sup>5)</sup>の手法に従っている。この手法は、全ての R 利用ユーザに共通した R パッケージのインストール環境を構築、管理する方法であり、定期的に環境の更新が行われている。

参考文献

- 1) A RISK-BASED APPROACH FOR ASSESSING R PACKAGE ACCURACY WITHIN A VALIDATED INFRASTRUCTURE  
<https://www.pharmar.org/white-paper/>
- 2) Case Study: R Package Risk Assessment at Novartis  
[https://github.com/pharmaR/case\\_studies/blob/main/Novartis\\_Rpkg\\_validation\\_2022](https://github.com/pharmaR/case_studies/blob/main/Novartis_Rpkg_validation_2022)

[0509.pdf](#)

- 3) Case Study: Risk Assessment of R Packages at Merck KGaA/EMD Serono  
[https://github.com/pharmaR/case\\_studies/blob/main/Merck\\_KGaA\\_Rpkg\\_validation\\_220202.pdf](https://github.com/pharmaR/case_studies/blob/main/Merck_KGaA_Rpkg_validation_220202.pdf)
- 4) External R Package Qualification Implementation at Merck: Case Study using GGally  
[https://github.com/pharmaR/case\\_studies/blob/main/External%20R%20Package%20Qualification%20Implementation%20at%20Merck.pdf](https://github.com/pharmaR/case_studies/blob/main/External%20R%20Package%20Qualification%20Implementation%20at%20Merck.pdf)
- 5) Shared Baselines  
<https://solutions.posit.co/envs-pkgs/environments/shared/>

### 3. R の教育研修

#### 3.1 R トレーニング事例

臨床開発における R ユーザの教育としては、これまで臨床開発領域で広く使われてきた SAS を利用してきた社員に対しての R 教育の事例がこれまでいくつか紹介されている（表 2）。事例の一つとして、一般に公開されているトレーニング教材を基に社内でカリキュラムを構築し、社内の経験者が講師となってトレーニングを行った事例がある<sup>1,2)</sup>。もう一つの事例として、ベンダーが提供するトレーニングパッケージを受講するという事例がある<sup>2)</sup>。両者に共通して、臨床開発における SAS プログラミングに精通した担当者を対象としており、一般的な受講者向けに設計されたオンライン R トレーニングコースとは異なり、受講者の日々の業務に関連したデータやシナリオを使用しつつ、SAS との違いにフォーカスするなどカスタマイズされている。それにより、受講者が共感しやすい身近な問題シナリオによって、効率的に学習を進めることができる内容になっている。また、受講者の不安解消や自発的な学習を促すため、受講者を少人数のコホートに分け、コホート内での協働を行っている、といった特徴がある。

表 2. 臨床開発プログラマ向けの R トレーニング実施事例

事例	内製トレーニング	外部トレーニング
概要	社内の R 経験者が講師として、4-6 名で構成されるコホートを指導する形式	5-6 名で構成されたコホートごとに e-learning システムの受講及びベンダー側の講師がサポートする形式
受講期間	3 ヶ月	6 ヶ月

講師	社内の R 経験者	ベンダー側の講師
対象者	SAS ユーザ	SAS ユーザ
コンテンツ	<ul style="list-style-type: none"> <li>● セッション前後の自己学習</li> <li>● 1 回 1 時間あたりの講師とのトレーニングセッション</li> </ul>	<ul style="list-style-type: none"> <li>● 1 コース当たり 1~3 時間前後の e-learning 動画視聴</li> <li>● 1 コース当たり 1~4 時間前後のセルフエクササイズ</li> <li>● ベンダー講師による Office Hour</li> </ul>
総学習時間	18~28 時間	25~30 時間

### 3.2 R トレーニングのカリキュラム

臨床開発において R を用いることを前提とした場合、R の基礎やパッケージ「[Tidyverse](#)」の習得を基本的な目標とした以下表 3 の内容についてカリキュラムに組み込むことが望ましい。Tidyverse はデータの読み込み、抽出、加工、可視化などのデータ分析における基本的な作業を効率的に行うことができるパッケージ群である。前述の Pharmaverse がリリースしている admiral についても Tidyverse をベースに作成されているため、Tidyverse の理解が重要である。また、表 3 の内容に加え、R のバージョン管理や R markdown <sup>脚注 1)</sup>、Shiny <sup>脚注 2)</sup> といった応用機能、パッケージ開発といった内容も有用である。

脚注 1) R markdown: R のコードで集計した結果(計算結果、グラフ、表など)を含めたレポートをマークアップ言語である Markdown 形式で作成するためのパッケージ。スクリプトとアウトプットを一つのファイルとしてまとめることができる。

脚注 2) Shiny: R 言語を用いて Web アプリケーションを作成するためのパッケージ。R での分析結果や分析のための BI ツール等を Web アプリケーションとして公開できる。

表 3. 臨床開発プログラマ向けの R トレーニングカリキュラム

カリキュラム	学習内容
R の基礎	<ul style="list-style-type: none"> <li>● R 言語概要</li> <li>● R 実行環境 (R Studio 等)</li> <li>● パッケージ</li> <li>● R のオブジェクト及びデータ型</li> <li>● 重要な演算子と関数</li> <li>● 欠損値</li> <li>● ループ・反復処理</li> <li>● Error/Warning</li> </ul>
Tidyverse の基礎	<ul style="list-style-type: none"> <li>● Tidyverse のパッケージ</li> </ul>

カリキュラム	学習内容
	<ul style="list-style-type: none"> <li>● データのインポート・エクスポート</li> <li>● 変数の作成・変更</li> <li>● データのサブセット化・ソート処理</li> <li>● データの要約</li> <li>● データの結合・転置</li> <li>● ggplot2 を用いた作図</li> </ul>
SDTM の作成	● テストデータを利用した SDTM 作成演習
ADaM の作成	● テストデータを利用した ADaM 作成演習
解析帳票の作成	● テストデータを利用した解析帳票作成演習

### 3.3 R 学習参考資料

R はデータ解析の分野で汎用的に用いられている言語であり、無償で公開されている資料も多く存在する。R トレーニング資料の作成や、受講者の自己学習に役立つと思われる公開資料について、表 4 にまとめた。講師の負担軽減や受講者の自発的・継続的な学習促進のため、このような公開資料を有効活用することを推奨する。

表 4. R 学習に関する公開資料

公開資料	概要
<a href="#">Dr. Garrett Grolemond's master the tidyverse workshop</a>	R for Data Science 執筆者の Dr. Garrett 氏の Tidyverse ワークショップ資料
<a href="#">Posit Videos</a>	Posit (旧 RStudio) 主催ウェビナーのアーカイブ
<a href="#">R for Clinical Study Reports and Submission</a>	R を用いて治験総括報告書を作成する方法についてまとめたサイト
<a href="#">Data Science Specialization</a>	Johns Hopkin University が提供するデータサイエンス講座
<a href="#">疫学や公衆衛生に関わる業務のための R</a>	Applied Epi が公開している疫学者や疫学実務者のための R ハンドブック
<a href="#">Pre-{admiral} Hackathon Workshop: Introduction to R for SAS Programmers</a>	CDISC が開催した SAS プログラマ向けの R 入門ワークショップの動画及び資料
<a href="#">SAS and R</a>	Bayer 社が公開している R のナレッジハブ (SAS プログラマ向け)

### 参考文献

- 1) Merck & Co., Inc. Jeff Cheng et al., PharmaSUG 2022 - Paper SI-055 "Implementing and Achieving Organizational R Training Objectives"  
<https://www.lexjansen.com/pharmasug/2022/SI/PharmaSUG-2022-SI-055.pdf>
- 2) Chugai Pharmaceutical Co., LTD. Taku Sakaue et al., PHUSE Japan Single Day Event 2022 "Introduction to an example of R usage"  
[https://phuse.s3.eu-central-1.amazonaws.com/Archive/2021/SDE/APAC/Virtual%20-%20Japan%20Winter/PRE\\_Japan04.pdf](https://phuse.s3.eu-central-1.amazonaws.com/Archive/2021/SDE/APAC/Virtual%20-%20Japan%20Winter/PRE_Japan04.pdf)

おわりに

OSS を承認申請に利用するための課題をタスクフォースメンバーで検討し、それらの課題を解決するための情報収集を行ったが、圧倒的に R に係る活動や文書が多く、製薬業界で承認申請に R を使っていこうとする、大きな流れを感じる機会となった。まだ、どの活動も過渡期にあり、今後、更に多くの R を使った承認申請事例や活用事例、R パッケージの開発が進んでいくことが予想される。今後のタスクフォース活動でも、引き続き OSS の活用事例や、OSS の管理について検討していきたい。

OSS を支えるものは、それを使うユーザのコミュニティであり、ユーザコミュニティが OSS の機能や品質の向上への大きな動力源となる。まずは、OSS を利用し、ユーザコミュニティの一員として積極的にフィードバックを続けることで、結果的に承認申請に利用できるようなものに成長する事を願っている。

日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2022 年度タスクフォース 5  
オープンソースソフトウェアの活用と留意事項チーム

菅波 秀規*	興和(株)
加藤 智子*	サノフィ(株)
土屋 悟*	住友ファーマ(株)
堀田 真一**	住友ファーマ(株)
坂上 拓**	中外製薬(株)
新井 賢太郎	ノバルティス ファーマ(株)
松永 友貴	ノバルティス ファーマ(株)
西村 力丸	ヤンセンファーマ(株)
生井 伴幸	中外製薬(株)
保田 昂之	中外製薬(株)

\* 担当副部長

\*\*推進委員