

2024年度 OSS利用状況アンケート報告書



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
2024年度 継続タスクフォース11

加藤智子、堀田真一、坂上拓、松永友貴、東島正堅、
西村力丸、生井伴幸、保田昂之、鈴木一平

データサイエンス部会 2024年度 継続タスクフォース11では、2022年度に引き続き、オープンソースソフトウェア（以下OSS）の利用状況についてのアンケートを実施した

➤ アンケート実施期間

■ 2024年9月17日（火）～ 2024年10月11日（金）

➤ 対象

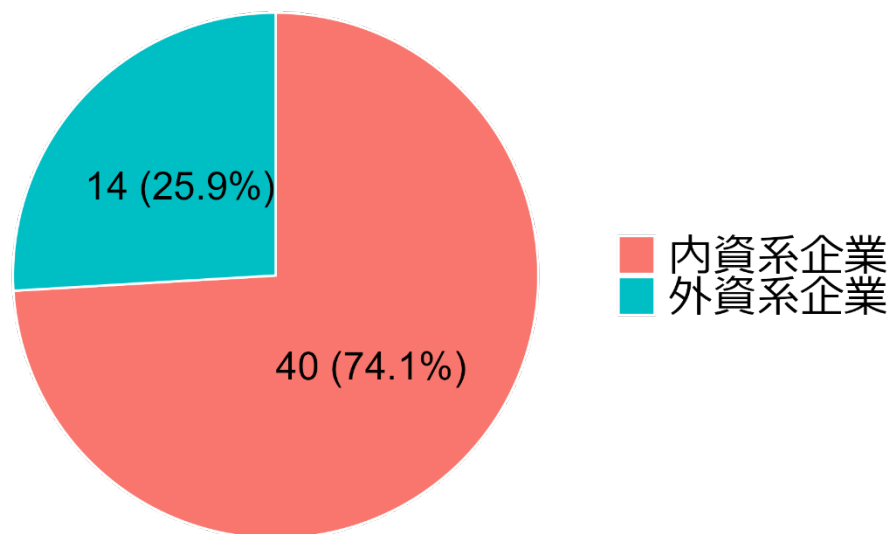
■ データサイエンス部会登録会社（内資系・外資系を問わない）

➤ 結果

■ データサイエンス部会登録会社67社中54社から回答を得た

■ 回答の集計結果と考察を次ページ以降にまとめた

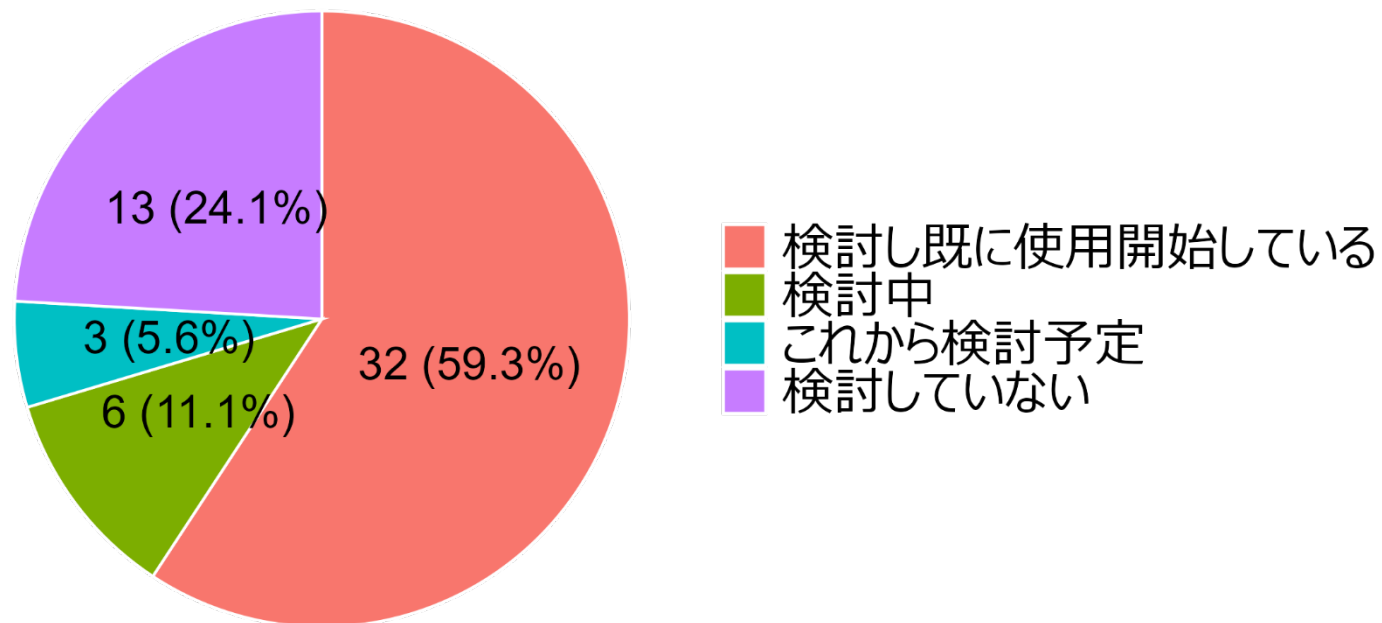
1. 企業形態を教えてください。(単一回答)(N=54)



考察

2022年度に実施したオープンソースソフトウェア(以下OSS)の利用状況のアンケートの回答数は55企業であり、ほぼ同じ回答数であった。

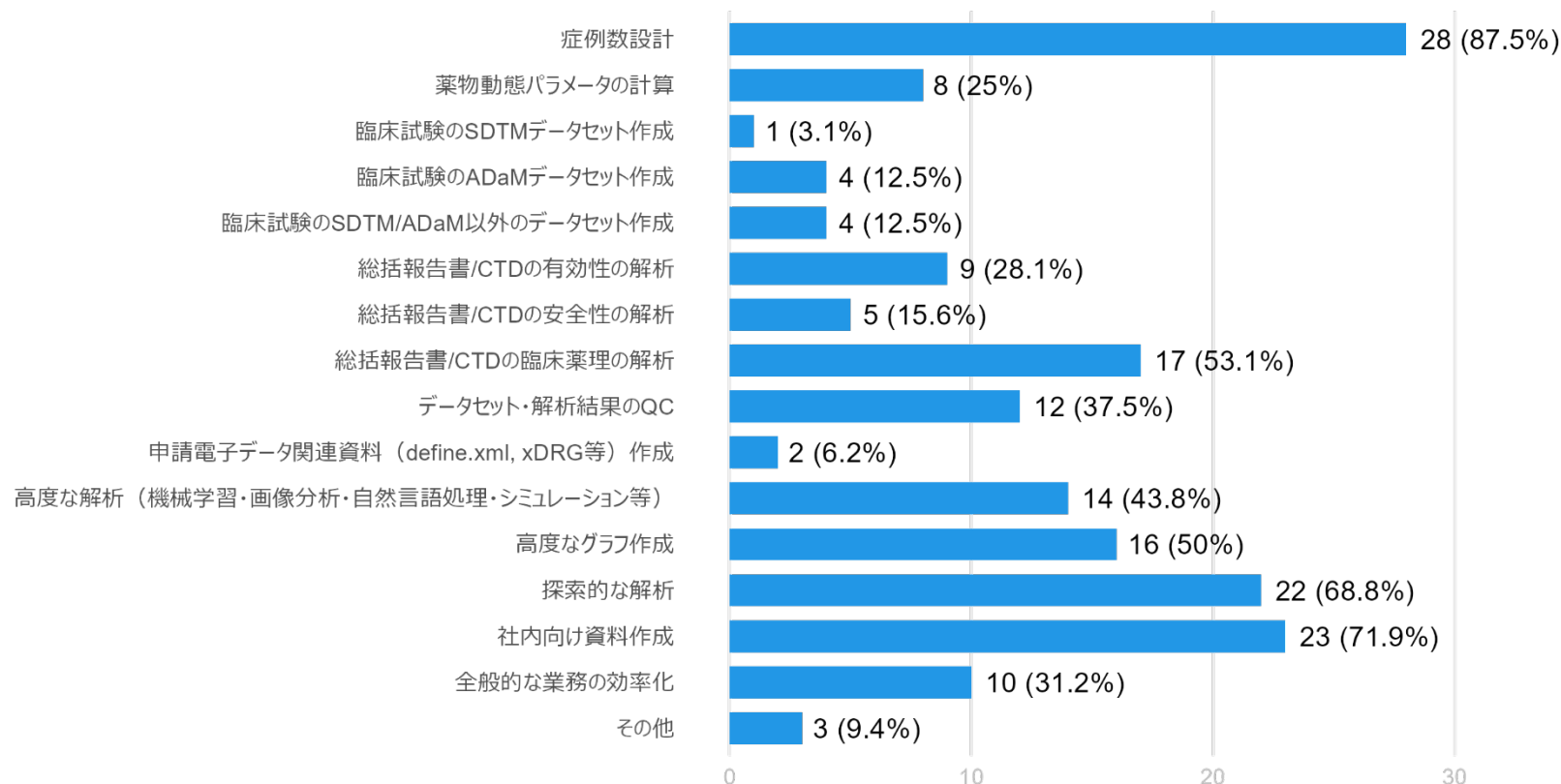
2. 臨床試験関連の業務で、オープンソースのプログラミング言語「R」の使用を開始または検討していますか？
(単一回答)(N=54)



考察

2022年度に実施したOSSの利用状況のアンケートの回答と比較して、「検討し既に使用開始している」及び「検討中」の回答数が微増した。依然として積極的にRを導入している企業とそうでない企業の温度差が感じられるものの、R利用についての関心は高まっていることが見受けられた。

3.臨床開発関連業務でRを「使用開始している」と回答された方への質問です。どの工程でR言語を使用していますか？(複数選択可)(N=32)



考察

2022年度のアンケートと同様に、症例数設計、探索的な解析、高度な解析、社内向け資料作成といった工程でのRの利用が多く、臨床試験のデータセットや解析帳票の作成とQC、申請電子データ関連資料の作成といった工程でのRの使用経験は少なかった。統計解析の領域では申請資料として直接提出しない資料の作成工程からRの導入が始まっていると考えられるが、一方で臨床薬理領域での解析におけるRについては事例が多かった。

4.前の質問で「高度な解析(機械学習・画像分析・自然言語処理・シミュレーション等)」と回答された方は、差し支えない範囲で、具体的な用途について教えてください。1/2ページ

教師あり学習(先行研究からの予測)

階層ベイズ法の実装検討

がん第I相試験におけるモデル支援型デザインの動作特性評価(シミュレーション)

オンコロジー領域の第I相試験でベイズ流の手法を用いる場合のシミュレーション

シミュレーションで複雑な乱数を発生させた場合に使用

試験デザインのシミュレーション

PK及びPK/PDシミュレーション

- ・薬物動態のシミュレーション
- ・モンテカルロシミュレーションによる検出力
- ・関数データ解析
- ・ランダムフォレスト/スパース推定
- ・CART
- ・クラスタリング
- ・BOINデザイン検討
- ・中間解析時のシミュレーション
- ・メタアナリシス
- ・Parallelパッケージを用いたBootstrapシミュレーション(並列計算)
- ・Sides法(エンリッチメント集団探索)
- ・ハザード比曲線(ノンパラメトリック, 平滑化)
- ・ニューラルネットワーク

4.前の質問で「高度な解析(機械学習・画像分析・自然言語処理・シミュレーション等)」と回答された方は、差し支えない範囲で、具体的な用途について教えてください。2/2ページ

- プレゼン資料としてR markdownの活用
- R shinyによるオンラインマニュアルの作成
- subgroup解析プログラム自動作成
- 試験計画段階における各種動作特性等を確認するためのシミュレーション
- 試験結果の解釈を深めるために探索的な位置づけとして実施する解析

探索的な解析の実施時に機械学習の手法を用いた。Oncology P1でシミュレーションを実施した。

B(H)LRMのシミュレーション

Graphical approach

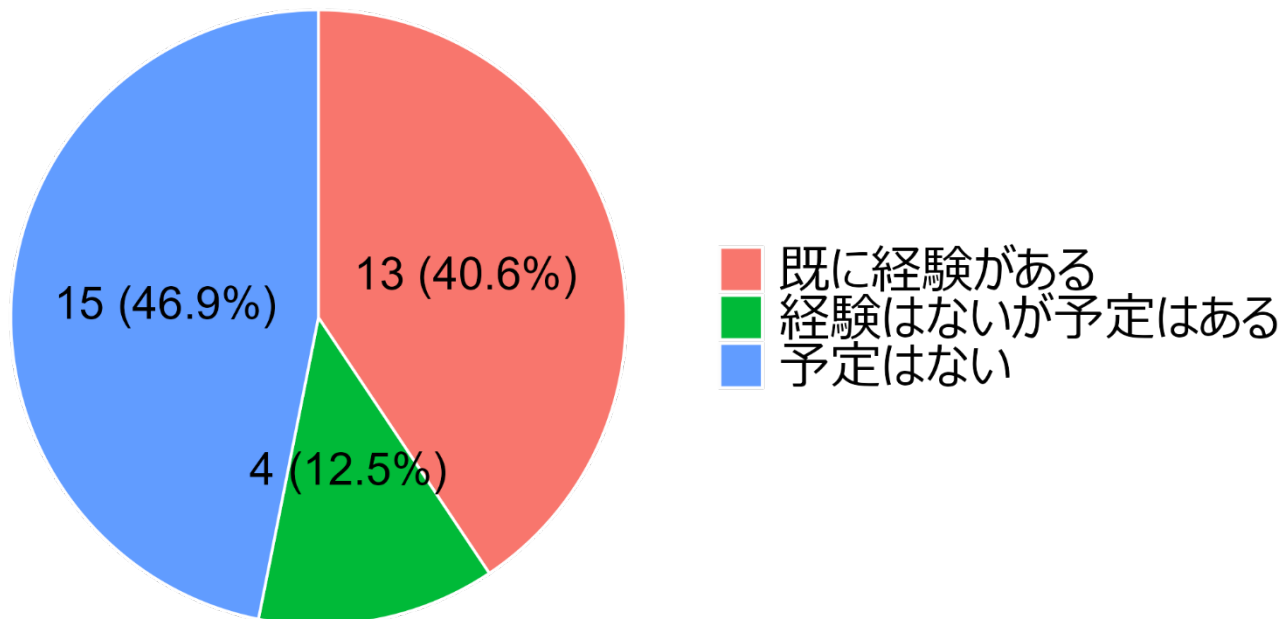
MCP-mod

PBPK

考察

モデリング&シミュレーションや機械学習を中心に、様々な利用事例が挙げられた。Rのパッケージのバリエーション(選択肢の多さ)や、Shinyによるアプリケーション化が容易であることなどがこの結果に繋がっていると推察できる。

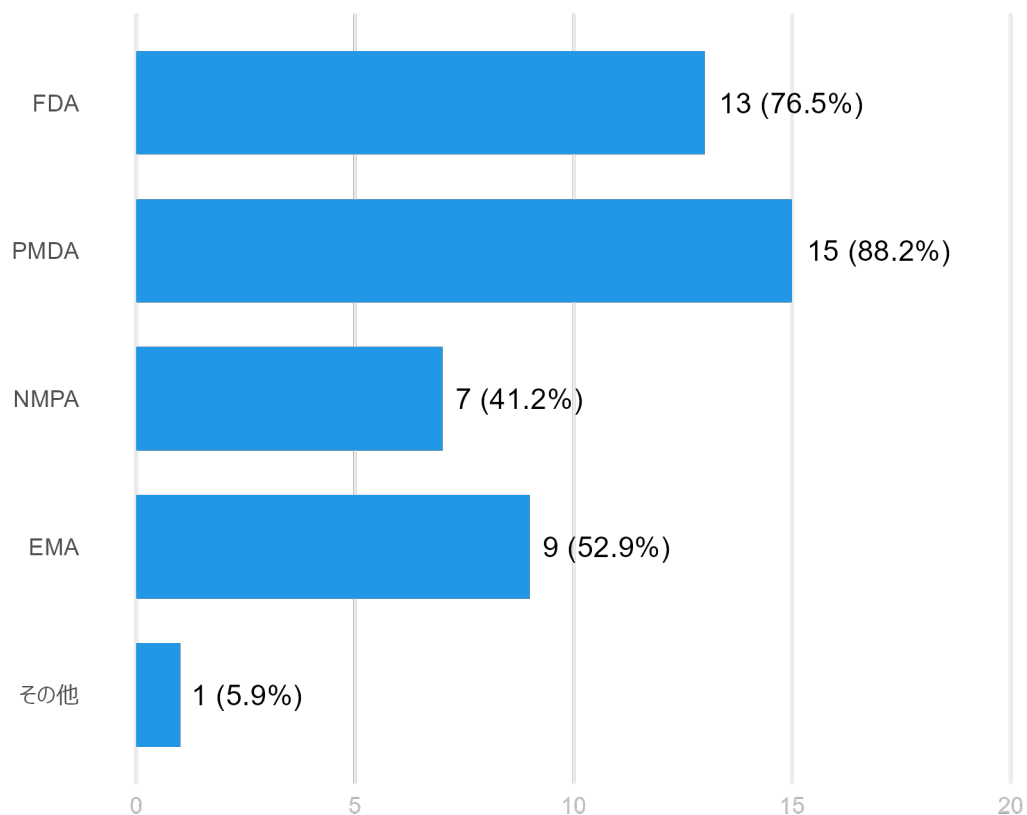
5. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。R言語を使用して作成した解析帳票やデータセット等の資料を規制当局に提出した経験または提出する予定ありますか？（単一回答）(N=32)



考察

2022年度のアンケートと経験の割合は大きく変わらないものの、「既に経験がある」と回答した企業が微減した。設問3の回答も踏まえると、提出経験がある企業においても多くの場合は臨床薬理領域の解析に関連する資料の提出経験に留まっていることが推察された。

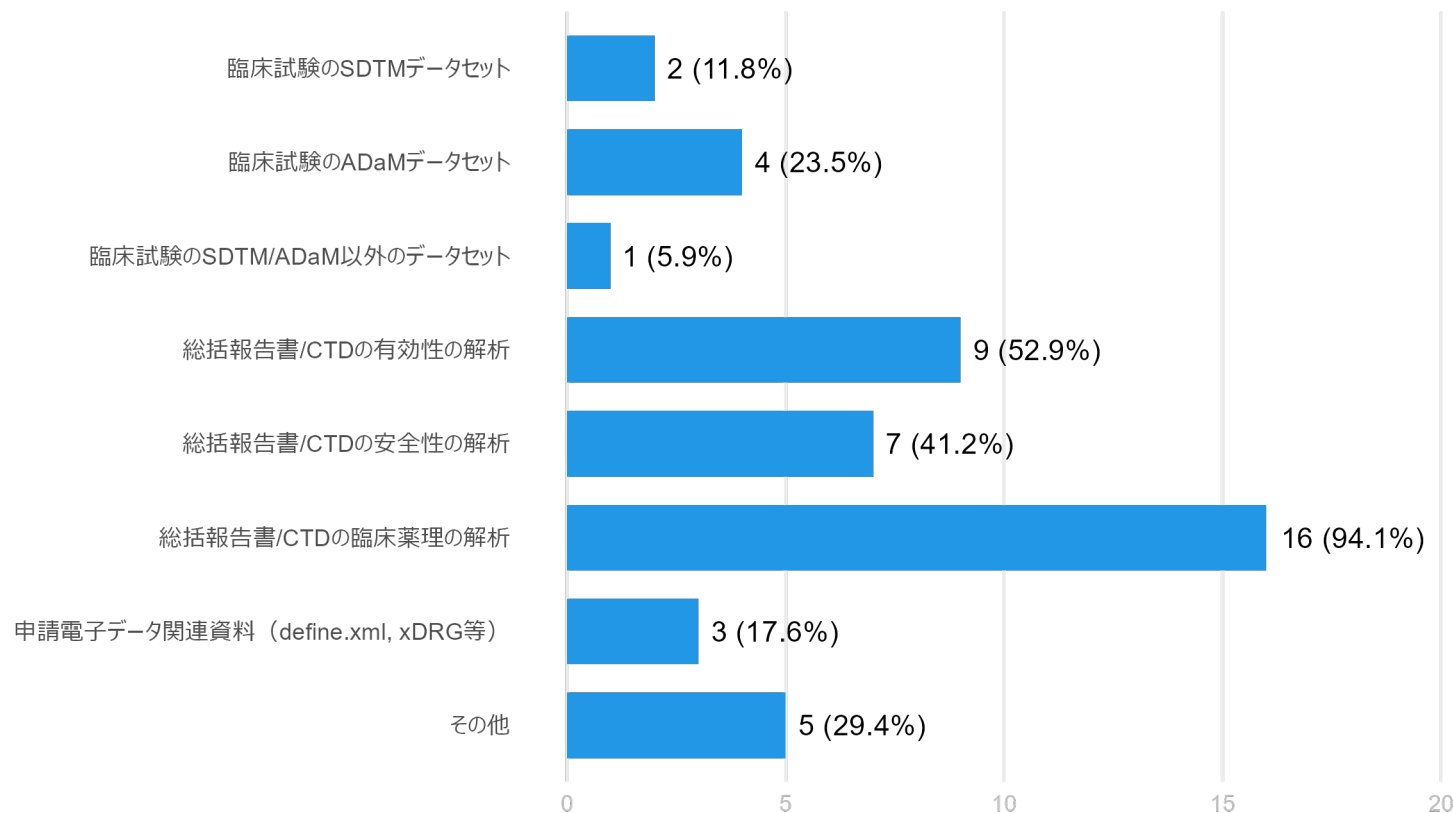
6. R言語を使用して作成した資料の規制当局への提出で「既に経験がある」または「経験はないが予定はある」と回答された方への質問です。どの規制当局への提出か、差し支えない範囲で教えてください。(複数選択可)
(N=17)



考察

FDA、PMDAへの資料の提出経験・予定がある企業が多かったが、他の当局でもR言語で作成された資料が受け入れられている傾向が示唆された。

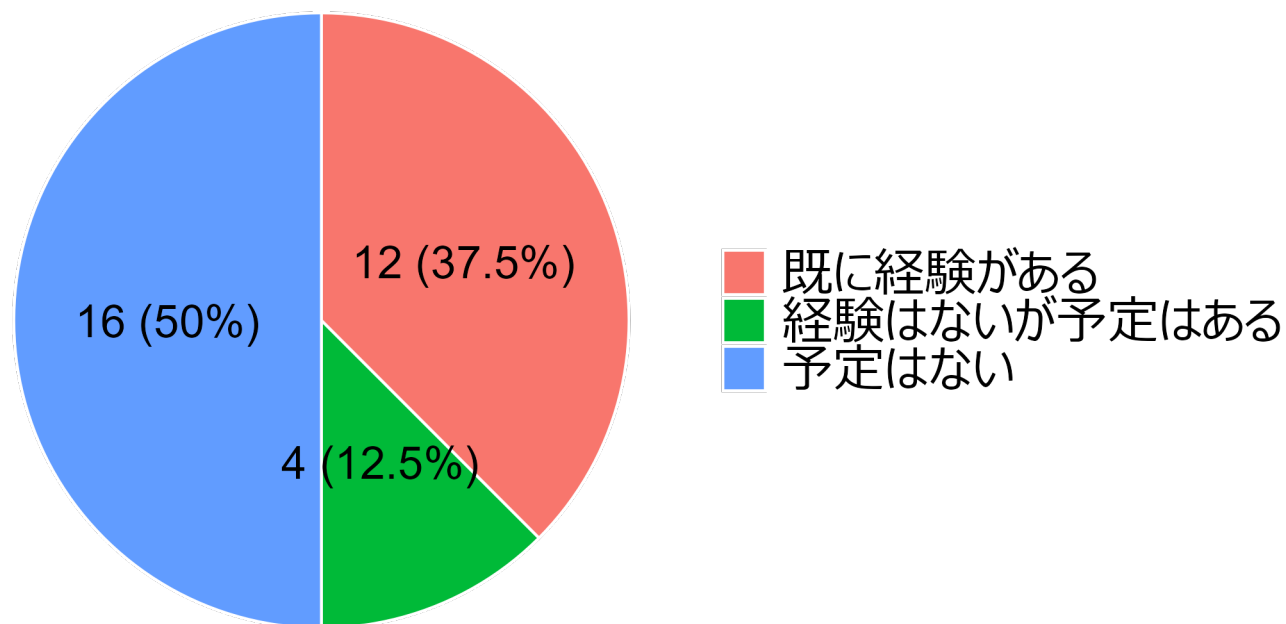
7. R言語を使用して作成した資料の規制当局への提出で「既に経験がある」または「経験はないが予定はある」と回答された方への質問です。どの資料をR言語で作成しているのか、差し支えない範囲で教えてください。(複数選択可) (N=17)



考察

臨床薬理領域をはじめとした総括報告書/CTDの解析に関連する資料の提出経験・予定が多かったが、R言語で作成したSDTM・ADaMの提出も行われている。

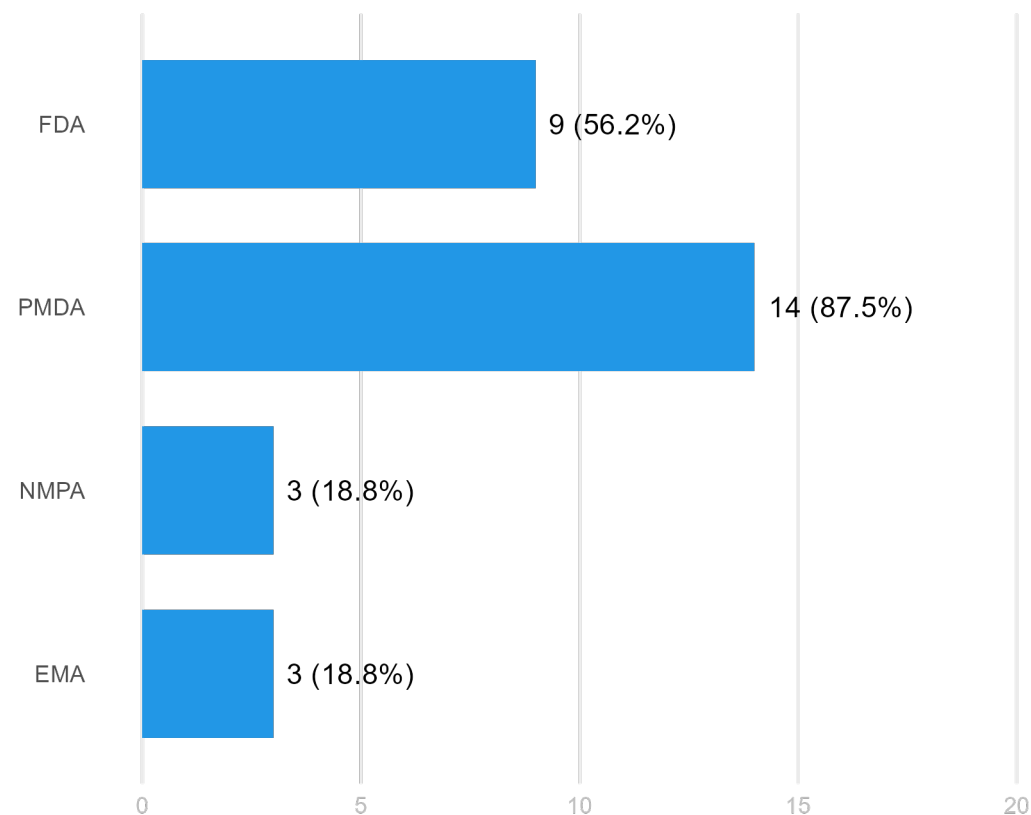
8. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。R言語で記載した解析プログラムを申請資料として提出した経験または提出する予定はありますか？（単一回答）（N=32）



考察

前回と同様の結果であったが、「既に経験がある」と回答した企業が微増した。

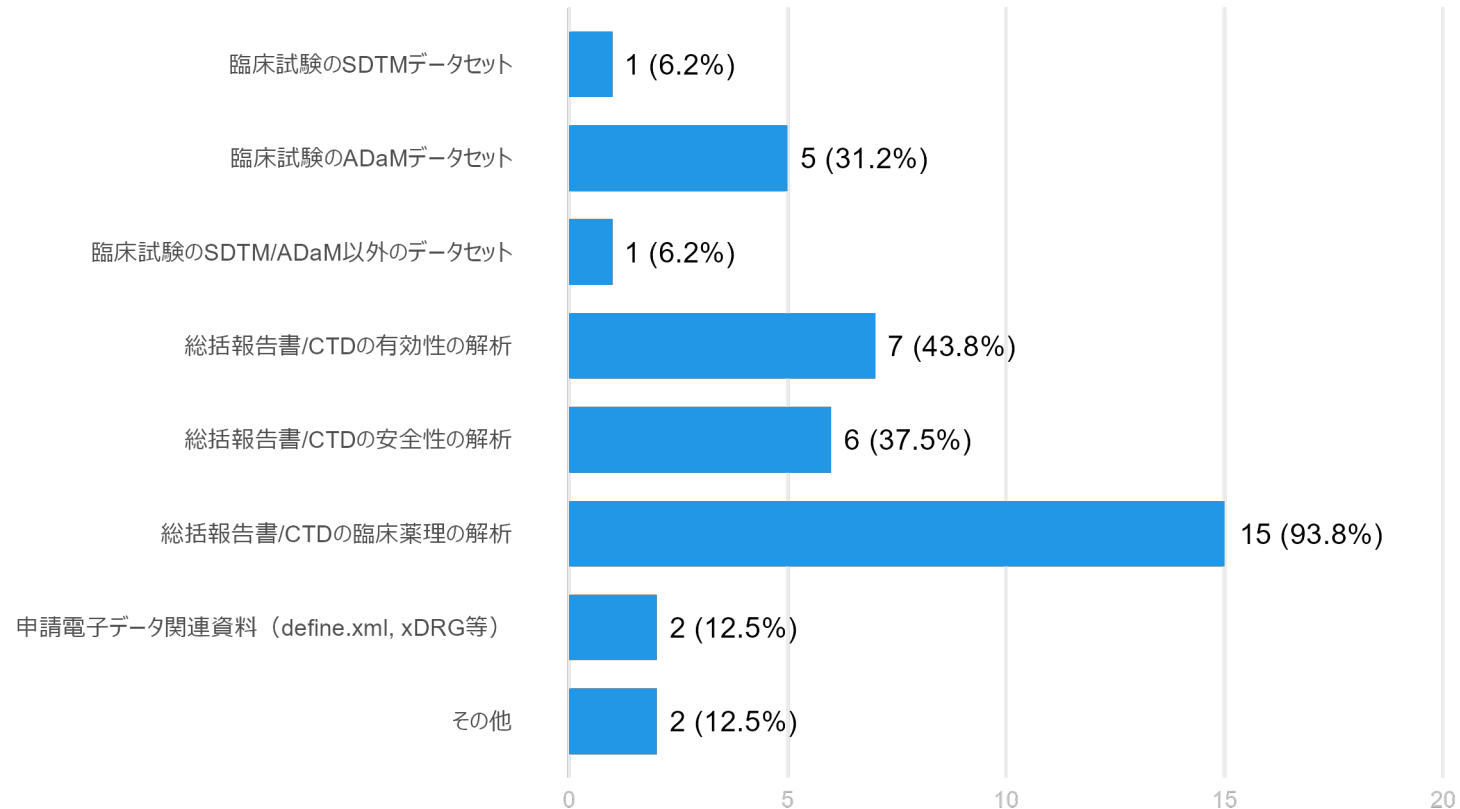
9. R言語で記載した解析プログラムの提出で「既に経験がある」または「経験はないが予定はある」と回答された方への質問です。どの規制当局への提出か、差し支えない範囲で教えてください。(複数選択可) (N=16)



考察

FDA、PMDAへの資料の提出経験・予定がある企業が多いという結果が得られた。

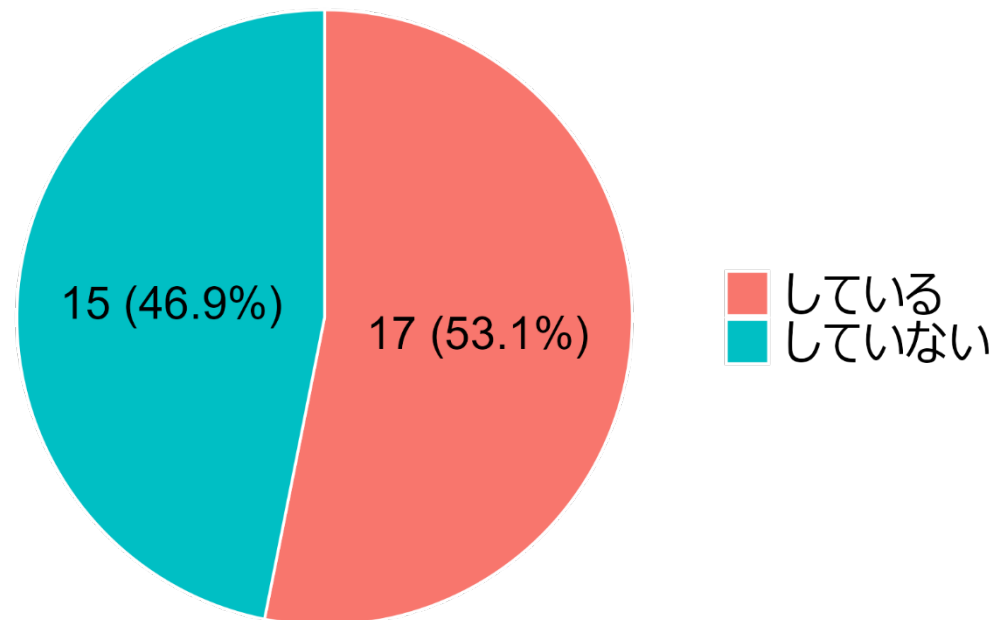
10. R言語で記載した解析プログラムの提出で「既に経験がある」または「経験はないが予定はある」と回答された方への質問です。どの資料の作成に使用しているRプログラムか、差し支えない範囲で教えてください。(複数選択可) (N=16)



考察

設問7の傾向と同様、臨床薬理領域の解析プログラムが多いが、SDTM・ADaM作成プログラムを提出している企業も見られた。

11. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。Rに対してバリデーションはしていますか？（単一回答）(N=32)



考察

Rの使用を開始している場合は、5割を超える企業でRに対してのバリデーションが実施されているという回答であった。

12. Rのバリデーションを「している」と回答された方への質問です。どのようなバリデーションを実施したか、差し支えない範囲で教えてください。

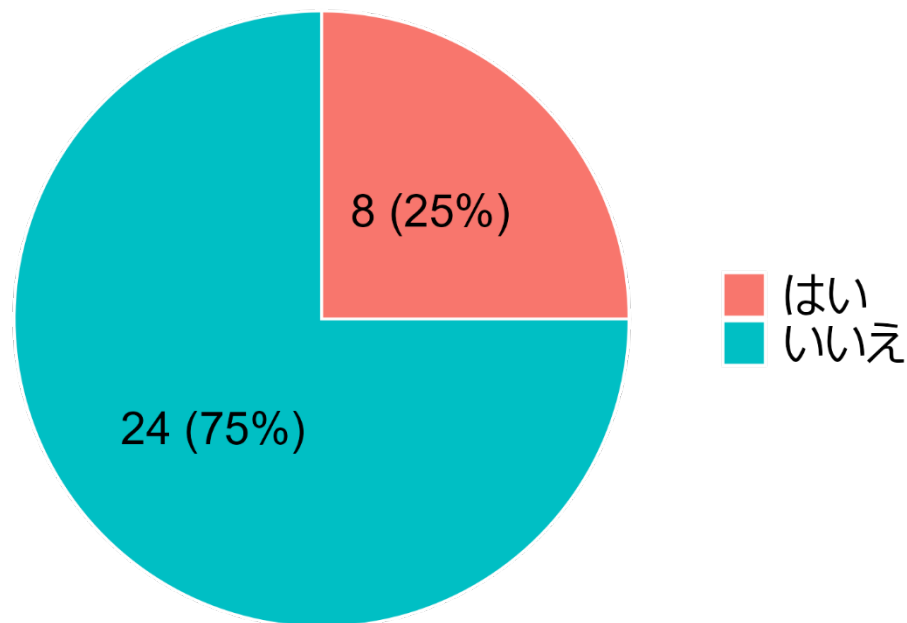


- SASによるダブルプログラミング - パッケージ/バージョンは、社内でバリデーション計画し、信頼性および正確性の検証が十分に行われ承認されたもののみ使用可能となる。
IT部門によるバリデーション、具体的なバリデーション方法は不明
R validation hubのWhite paperに従って "A Risk-based Approach for Assessing R package Accuracy within a Validated Infrastructure"
Rのインストールが手順通り行われたかの確認した。R studio上でグラフが正常に表示されるか確認した。市販のソフトウェアで一部の機能でRを使っている場合も動作確認をした。
Rの各packageのversionを固定してバリデーションを実施
SASでのoutputとの結果の一致の確認
SASを使って同じ結果が得られることを確認した
TranscelerateのMSAに準じてバリデーションを実施。
ダウンロード数やjournalでの引用数等をもとに、パッケージ毎にリスク評価を行っている。
ダブルプログラミング又はコードチェック
テストデータを用いた解析結果の検証（動作確認）
バリデーションされたR製品を利用及び多言語によるダブルプログラミングの実施。
海外主導で行っているため不明
当社の統計コンピューティング環境は GxP検証済みであり、Rパッケージの取り込みに関しては R検証ハブの推奨事項に従っています。またRパッケージを評価するためのメトリックを作成するために、RパッケージのRiskmetricなどを使用しています。

考察

Rを使用するにあたっては、企業により様々な方法でRに対するバリデーション、R言語を使用した結果に対するバリデーションを実施しているという結果であった。

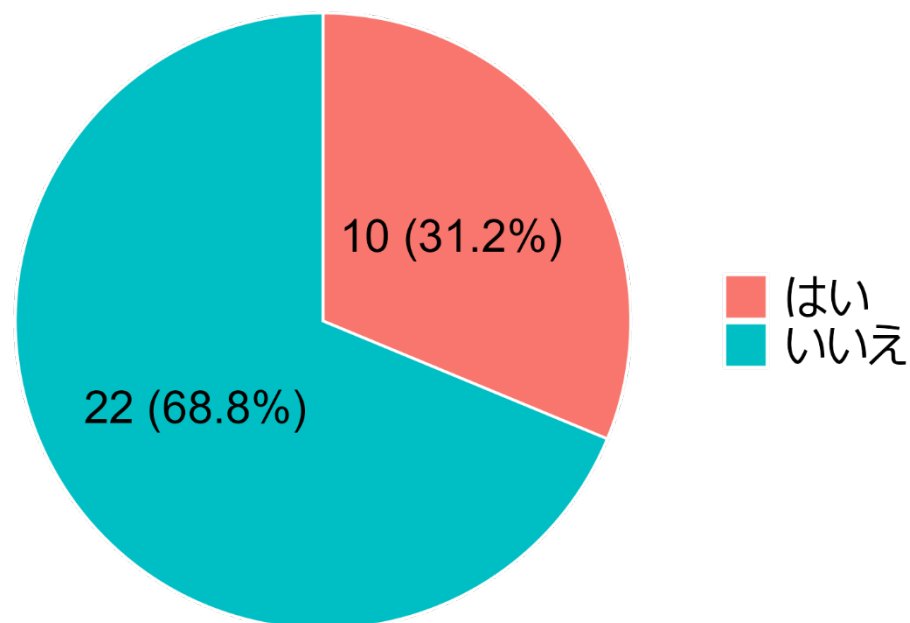
13. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。Pharmaverseが作成したパッケージ(admiral、rtables、pkglite等)を使用していますか？(単一回答)(N=32)



考察

Pharmaverseが作成したパッケージを使用している企業は、1/4という結果であった。設問3の「どの工程でR言語を使用していますか？」の回答にあるように、プログラムの提出を規制当局に要求されない工程で使用している企業が多いことが背景にあると考えられる。

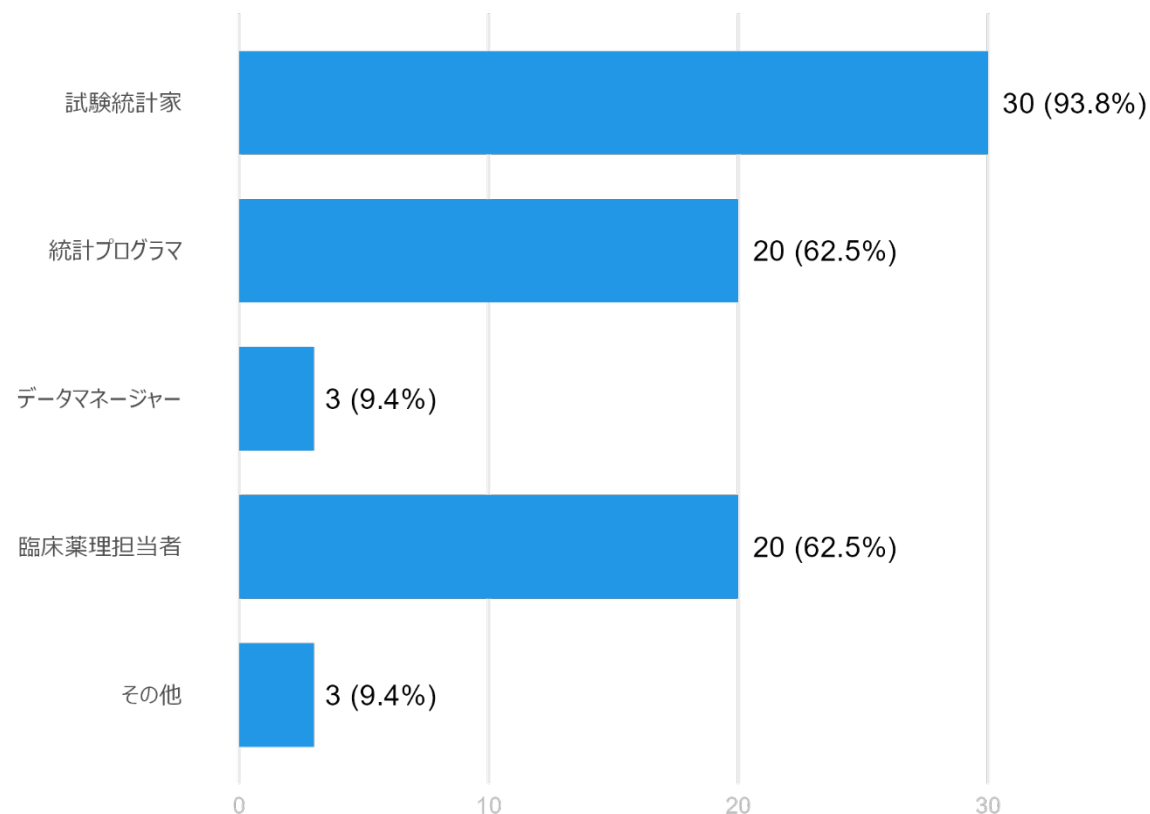
14. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。Rプログラムでの解析業務を外注した経験はありますか？（単一回答）(N=32)



考察

Rの使用を開始している企業のうち、約3割でRプログラムでの解析業務を外注した経験があるとの回答であった。2022年度に実施したアンケートでは、約1割(N=55)の企業で外注経験があるとの結果であり、外注の経験がある企業が増加したという結果が得られた。

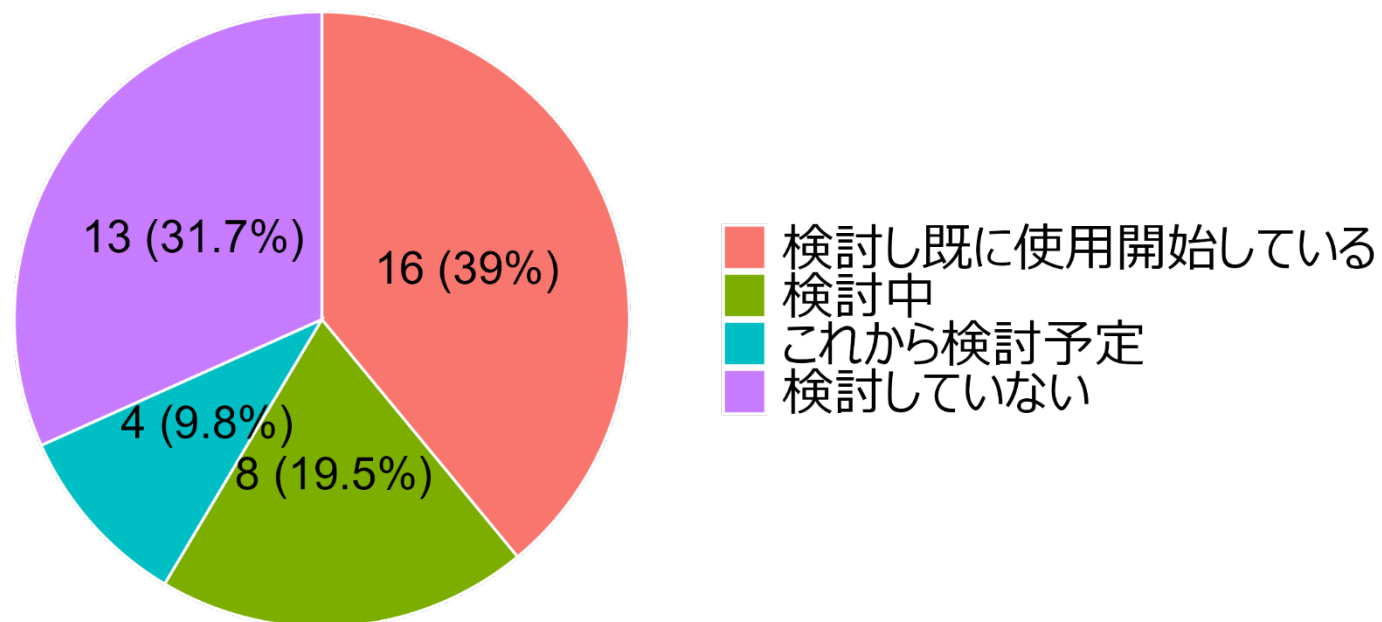
15. 臨床開発関連業務でRを「使用開始している」と回答された方への質問です。臨床開発関連の業務でR言語を利用している担当者を以下から選んで下さい。(複数選択可)(N=32)



考察

試験統計家、統計プログラマ、臨床薬理担当者において、R言語の利用者が多いという結果であった。設問3の「どの工程でR言語を使用していますか？」に対して回答を得た工程に携わる担当者での利用となっていた。

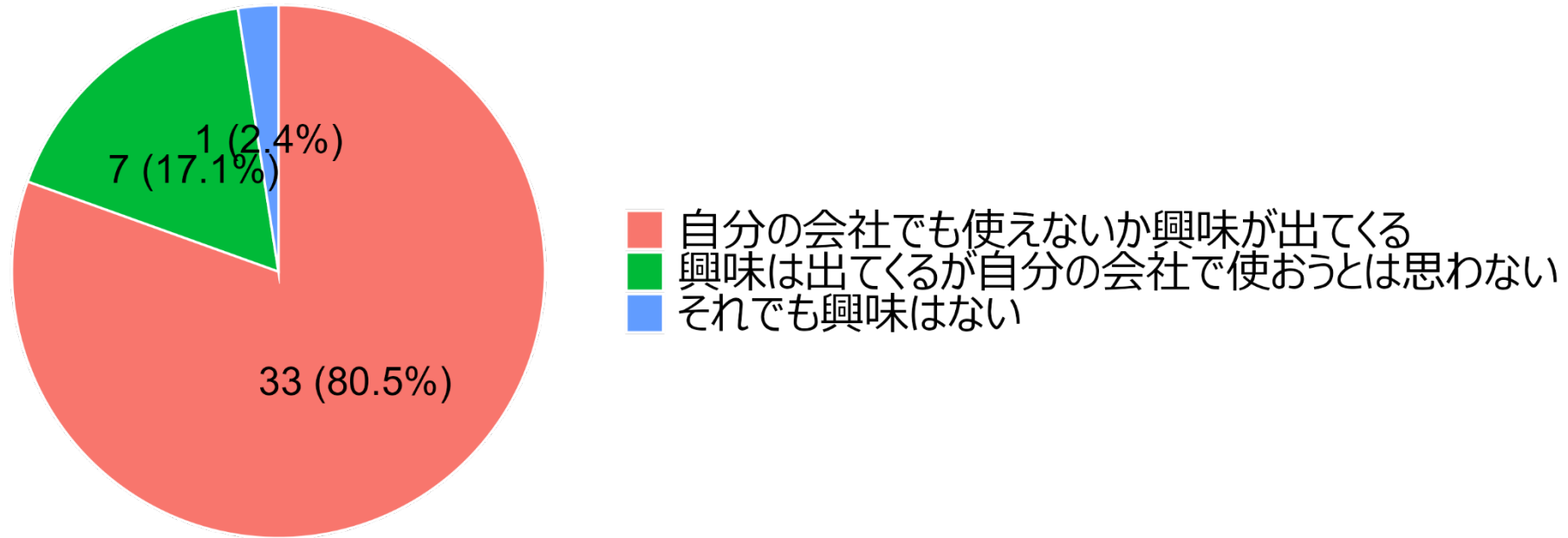
16. 臨床開発関連業務でRの使用を「既に開始」「検討中」「検討予定」と回答された方への質問です。臨床試験関連の業務で、RでインタラクティブなWebアプリを作ることができる機能「R-shiny」を使用を開始または検討していますか？（単一回答）（N=41）



考察

設問2でRの使用を「既に開始」「検討中」「検討予定」と回答した41社のうち、16社(39.0%)が既にR-shinyを使用開始しているという結果であった。「検討し既に使用開始している」「検討中」「これから検討予定」と回答した企業の総数は28社(68.3%)であり、R-shinyによるWebアプリの作成への関心の高さがうかがえる。

17.臨床開発関連業務でRの使用を「既に開始」「検討中」「検討予定」と回答された方への質問です。R ConsortiumではFDAのPilot申請制度を利用して、Rで作成した申請電子データや解析結果を使ってFDAで審査できるかを検証しています。4回目のPilot申請では、規制当局がWeb画面上でパラメータを変えることで追加解析を行えるR-shinyプログラムを提出しようとしています (<https://www.appsilon.com/post/testing-containers-and-webassembly-in-submissions-to-the-fda>)。将来、このようなプログラムを提出することで規制当局との申請後のディスカッションがスムーズになるとしたら、R-shinyに興味が出ますか？(単一回答)(N=41)



考察

設問2でRの使用を「既に開始」「検討中」「検討予定」と回答した41社のうち、33社(80.5%)が規制当局へのR-shinyプログラムの提出について「自分の会社でも使えないか興味が出てくる」という回答であった。多くの企業が申請後のディスカッションをスムーズにする目的での、R-Shinyの承認申請への利活用可能性に関心を示していることがわかった。

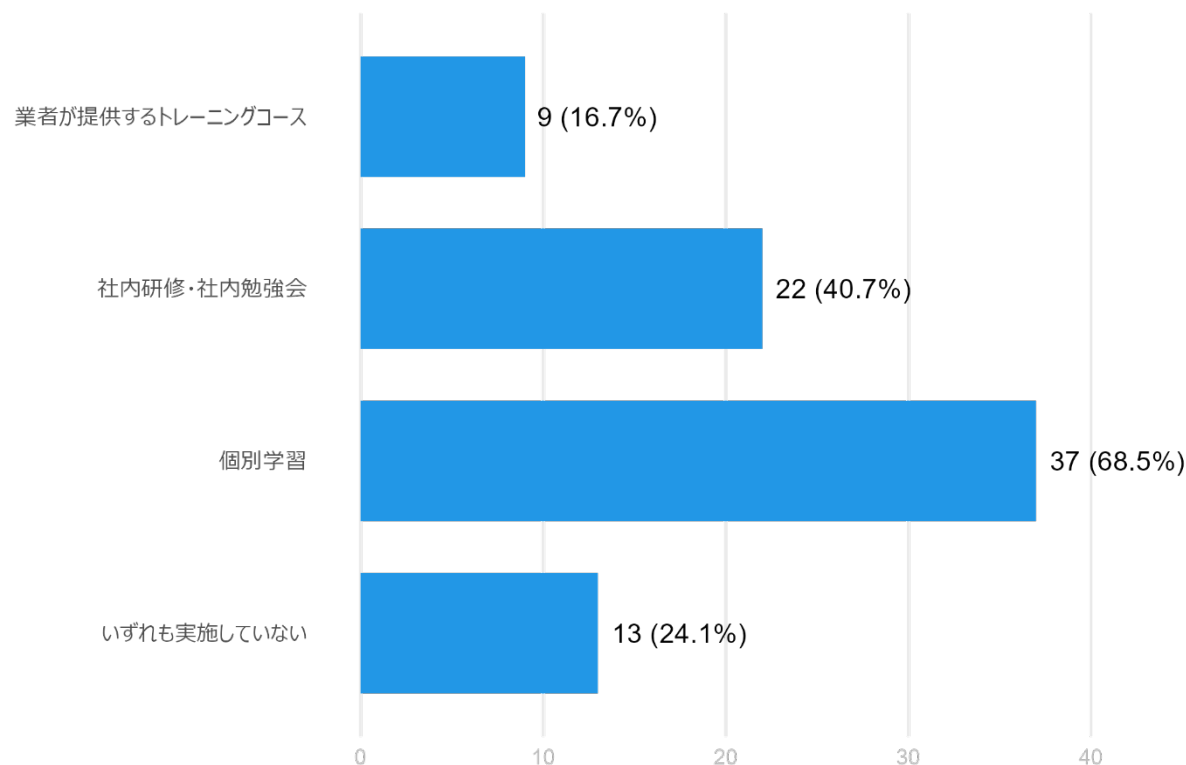
18. 臨床開発関連業務でRを「使用開始している」と回答されなかった方への質問です。臨床試験関連の業務でRを使おうとすると、障害になることは何でしょうか？差し支えない範囲で教えてください。

SASの使用で問題ないため、Rの導入は検討していない
信頼できない
Rで算出された解析結果のバリデーション
sasで業務を遂行できてしまうため。セキュリティの問題。
Rでないとできない業務が多くなり使いこなせるメンバーが少ない
社内セキュリティの関係で「R」パッケージの利用が出来ないため
社内の体制、当局の受け入れ
バリデーションが懸念点と認識しています。
障害になることではないのですが、業務上はSASを利用していることもあり、Rの業務での使用を検討していないこと。
(少なくとも弊社内では)バリデーション方法が整備されていない。また、トレーニング方法が確立していないため担当者の作業品質に懸念がある。
・Rを使った解析結果を規制当局がすんなり受け入れるのかの疑問がある。
・SASで解析した場合よりも、解析プログラムの妥当性(Rの計算ロジックの妥当性)の当局への説明が必要になるような印象がある(経験が無いので、想像となります)。
・CDISC(申請電子データ)への対応。
・SASとRの2つのソフトウェアに関して知識を習得する必要がでてくる。

考察

Rを臨床試験関連業務に導入する障害として、2022年度に実施したアンケートと同様、「SASで十分である」「バリデーションに対する懸念」「社内の体制やセキュリティ」「Rの知識習得にかかるトレーニングコスト」「当局の受け入れに関する懸念」といった回答が挙げられた。

19. 全員への質問です。社内で実施しているRのトレーニング/学習状況を教えてください。(複数選択可) (N=54)



考察

2022年度実施のアンケートからの変動は、「業者が提供するトレーニングコース」は12.7%→16.7%、「社内研修・社内勉強会」は25.5%→40.7%、「個別学習」は60.0%→68.5%、「いずれも実施していない」は34.5%→24.1%であった。何らかのトレーニングを実施する企業が増え、なかでも「社内研修・社内勉強会」を実施する企業が大きく増加したことがわかる。

20. Rの何らかトレーニング・学習を実施していると回答された方への質問です。差し支えない範囲で、学習を進める上で課題があれば教えてください。(1/2)



公開されているRトレーニングの種類が多くて、どのトレーニングが自分たちに最適かどうか個人で判断するのが難しい
社内トレーニングの資料がまだ十分には整っていない
・SASよりも中身がブラックボックス(SASではhelpが利用できるが、Rのマニュアルのみでは理解が難しい) ・CSV/バリデーションが取れているか否かと業務で利用できるか否かが直結する。バリデーションが取れたvalid-Rなどを使用する場合は費用が掛かるため、Rのお手軽感が薄れる ・使用環境によっては定期的なRのバージョンアップが難しく、使用するパッケージの推奨バージョンとの乖離が生じることがある ・似たような機能のパッケージがあり、どのパッケージを使うべきか判断に迷う(当局提示用の解析の場合、推奨パッケージなどがあれば嬉しい)
・分からないことが多すぎて調べることばかりになる(精通している方からの勉強会でクリアできている / AIに質問することでGoogle検索よりも効率的に理解できる) ・どのパッケージ、技術が将来的にも使われ続けるか、判断が難しい(例えば、Pharmaverseであったとしても、より便利なパッケージが出てきてしまっ廃れてしまう...という場合、学習した内容が使えなくなってしまうリスクについて、オープンソースソフトウェアであり、誰しもが開発出来てしまうが故に、そのようなリスクはSASよりも大きいと想像されました。)
SASをメインに使用しているため、Rの学習のための時間を確保することが難しい。
SASでの実業務が優先されてしまう(実業務でのRの必要性が高くない認識)
実務での使用がなく経験に基づいたトレーニングや学習ができない
社内研修を実施したが、学習を進める上の課題は特に無かった。
業者のトレーニングコースだとコンテンツが英語かつ、有料が多い。バージョンによって結果が違う等、Rパッケージのバージョン管理が難しい。他の解析ソフトウェアの結果と異なる際に困ることがある。
社内研修資料は頻用されるデータ加工テクニック等入門的な内容のみで、その後の発展的な学習は、個人の趣味や業務で直面する課題にあわせ担当者任せになっている。解析担当として目指す姿とそれに合わせたRの学習ガイドがあるとよいが、パッケージが日々更新・新規追加される環境下でどう対処すべきか悩ましく思う。
社内でのパッケージ利用方針(社内標準パッケージや公開パッケージの利用のルール等)のトレーニングは必要だと考える。
ユーザー数が多い場合はタイムリーなサポートが難しい。また、すべてのユーザーがパッケージ開発のスペシャリストである必要はないので、ユーザーの目的に合わせたトレーニングを準備し、受講する必要がある。
製薬企業向けの日本語のトレーニングが無い。
社内トレーニングだけでは不十分で、個々の自主的な学習が必須だと感じる。また、学んだことをすぐ実践に移せる環境を整えることがスキルの定着のためには望ましい。

20. Rの何らかトレーニング・学習を実施していると回答された方への質問です。差し支えない範囲で、学習を進める上で課題があれば教えてください。(2/2)



追加でパッケージをインストールする場合、社内のシステム管理者へ許可が必要(セキュリティ上、自動でパッケージ追加できない)
臨床開発に関連するRの文献がまだ少なく感じる。R未経験の方と経験のある方とでは理解や学習の進度に差が生じ、組織的な学習を進めると経験者の負担になる。
学習したものの、実際に解析業務にまだ使う段階にはなっていないため、学習したことを忘れてしまう。解析を目的としたR packageに関する日本語コンテンツが少ない。Rと共に、Rを実行する環境にかかわる知識を要する事があり学習範囲が広い。
トレーニングや勉強会を行っても使わないとすぐに忘れてしまう。Pharmaverseのパッケージはまだ発展途上であり、しばらく時間を置くと引数などが増減してしまうため再度勉強が必要。
他の業務が忙しい。パッケージ次第だが、使えるようになると便利だとは感じている。
<ul style="list-style-type: none">・自身の書いたプログラムが簡素化できているか、判読性があるかがわからない・新しいRのパッケージをフォローできていないかもという不安・ネット情報にデマ(誤り)も多い点・Rに組み込まれている検定のp値と共に出てくる信頼区間が対応していないといった細かな問題に気がつきづらい
Rに限った話しでは無いが、継続的な利用機会と学習機会が整わなければ定着しないため、Rを使う必要性が低い現段階では個人の意欲に左右されてしまう。
Rに精通した経験者が社内に不足しており、研修企画においては外部機関に依存する傾向がある点。
大学で使用していた人が学習しているのみ
現在は個々のスキルやモチベーションに依存している状況であり、人材育成の点から、系統的なトレーニングの実施が課題(適切なトレーニング資材やトレーナーの確保を含む)
SASのような自由度の高い帳票作りをRではできないことと、Rは日本語に弱く日本語版データセットの読み書きができないことが障害であり、解決が課題となっております。
取り組む時間の捻出など

考察

Rの学習を進める上での課題として、「日本語のトレーニングコンテンツの少なさ」「トレーニング体制の検討やトレーニング資材の選定」「Rやパッケージのバージョン管理方法の検討」「パッケージの更新や新たなパッケージに対する継続的かつタイムリーな学習機会の構築」「実業務への利用機会の少なさ」「学習時間の確保」といった回答が多く挙がった。