

Single Day Event

Quality assurance of Open Source Software (OSS) in statistical analysis of clinical study data for drug development

[R Quality Assurance Team]

Yoshiumi Ouchi (Novartis Pharma K.K.)

Hisae Watanabe (ONO PHARMACEUTICAL CO., LTD.)

Masataka Sano (Takushoku University)



Disclaimer

- The views and opinions expressed here are those of the presenters and should not be construed as representing the views and policies of Novartis



Agenda

- Definitions of Terms
- Background
- Objectives
- Survey on the status of use of statistical analysis software and quality assurance
- R Quality Assurance Opinion Exchange Meeting
- Summary and future prospects

Definitions of Terms

- Validation
 - Confirmation by objective evidence that product, services, processes, or systems meet needs for their intended use or application
Note: Objective evidence includes the results of experiments, tests, simulations, etc.
- Verification
 - Confirmation by objective evidence that a product, service, process, or system fulfills specified requirements
Note: A specified requirement is a requirement that is explicitly stated in document
- Quality Assurance
 - Systematic activities carried out by an organization to ensure, confirm, and demonstrate that the needs of customers and society are met



Background

- OSS
 - In the statistical analysis of clinical studies in drug development, OSS, especially the use of R language and R package, is drawing attention.
- Quality Assurance Issues
 - It is known that there are issues in validation and verification from the viewpoint of ensuring the reliability required by ICH E9.

“Quality Assurance issues” is one of the hurdles for the use of OSS.

Objectives

- The followings were conducted to understand the actual status of use of OSS and quality assurance for OSS in the industry
 - Survey on the status of use of statistical analysis software and quality assurance
 - Opinion exchange meeting on the use of OSS and quality assurance by experts on the use of OSS

Information from these two activities is presented in this presentation.
In addition, consider the future prospects for the use of OSS
in the industry based on the obtained information.

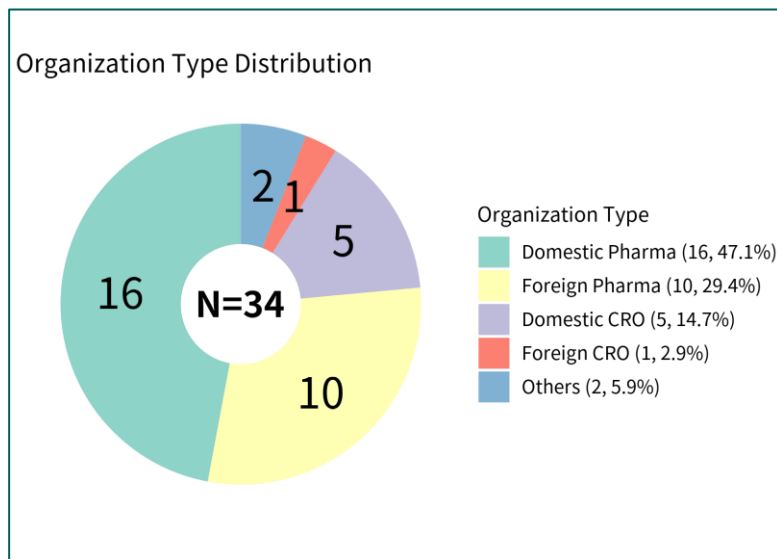
Survey on the status of use of statistical analysis software and quality assurance

*hereinafter referred to as "Survey"



Survey

- Target: CDISC Japan User Group ADaM
 - Approximately 60 members at the time of survey
- Period : 2025/8/22 - 2025/9/22
- Num. of responses: 34 (organizations)
 - Multiple participants from the same organization, a representative responds
- Items
 - See Appendix





Summary of Survey Results

- Statistical analysis software currently used
 - “Commercial Only (SAS etc.)” was most used, followed by “Commercial software + OSS (SAS & OSS)”.
- Currently used OSS
 - “R” was most used, followed by “Python”.
- R analysis environment and validation method for R packages
 - An equal number of “”, “FDA CSA Quality Assurance*” and “Not Implemented”.

*Draft Guidance on Computer Software Assurance for Production and Quality System Software
(at the timing of survey, guidance was draft version)



Summary of Survey Results

- Statistical analysis software expected to be used in 1 to 2 years
 - “Commercial + OSS (SAS Main) ” was most common, and “Commercial Only (SAS etc.)” was not frequent.
- OSS expected to be used in 1-2 years
 - “R” was most common, followed by “ Python.”
- Plans for validation and quality assurance activities for R analysis environment and packages
 - The most common response was “Under consideration, ” followed by “Own QA method” and “FDA CSA Quality Assurance”.
Some organizations purchased R packages for which quality control was performed for a fee.



Summary of Survey Results

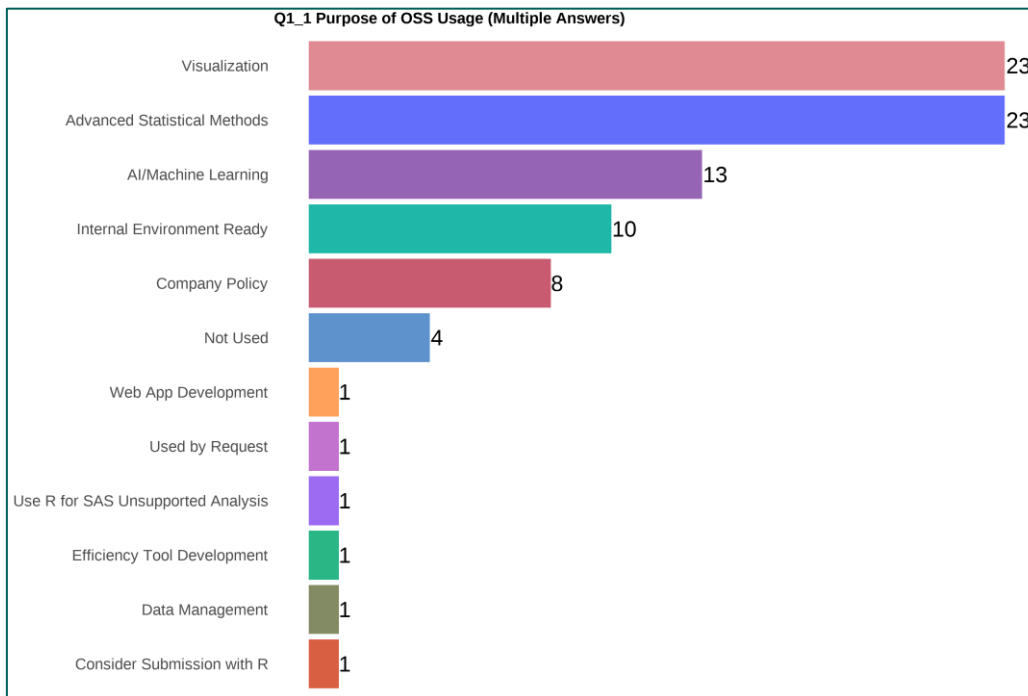
- Expectations for the use of OSS
 - “Use of cutting-edge technology” and “Cost reduction” were most common with the same number, followed by “Standardization” and “AI/Machine Learning”.
- Issues related to the use of OSS
 - “Lack of internal understanding/training” and “Challenges in validation/quality assurance” of OSS were the most common with equal number, and about 80% of the respondents considered them to be issues.
Followed by “Concerns about regulatory compliance” and “Cost/resource constraints”.



Survey Q1

Is OSS currently used internally? If yes, tell us the main purpose of use. (Select all that apply)

*Only this question could be answered including analyses for the clinical study report and analyses other than those for the CTD.

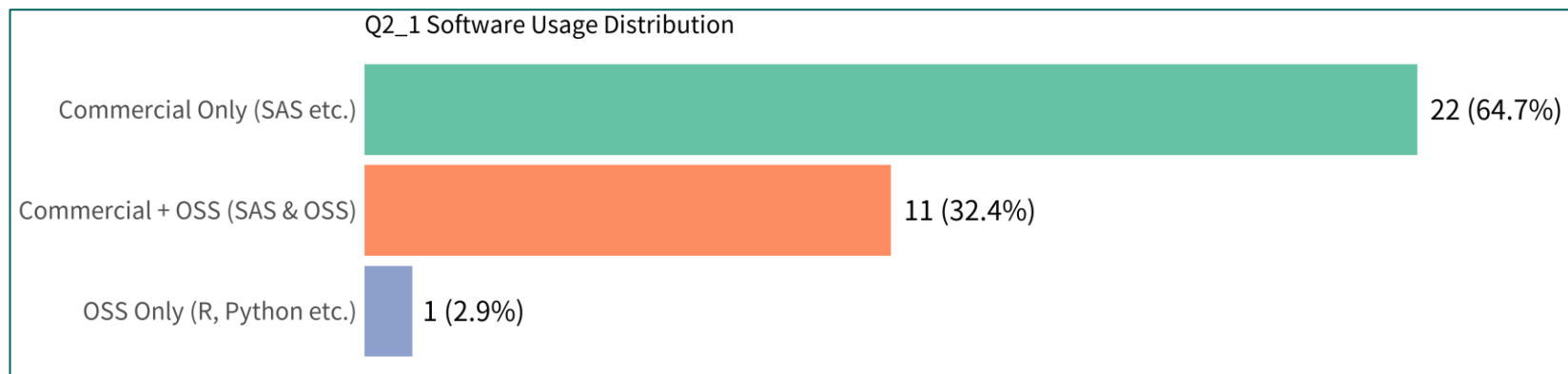


Q2 Current statistical analysis operations (analysis for the clinical study report and analysis for the CTD)



Survey Q 2.1

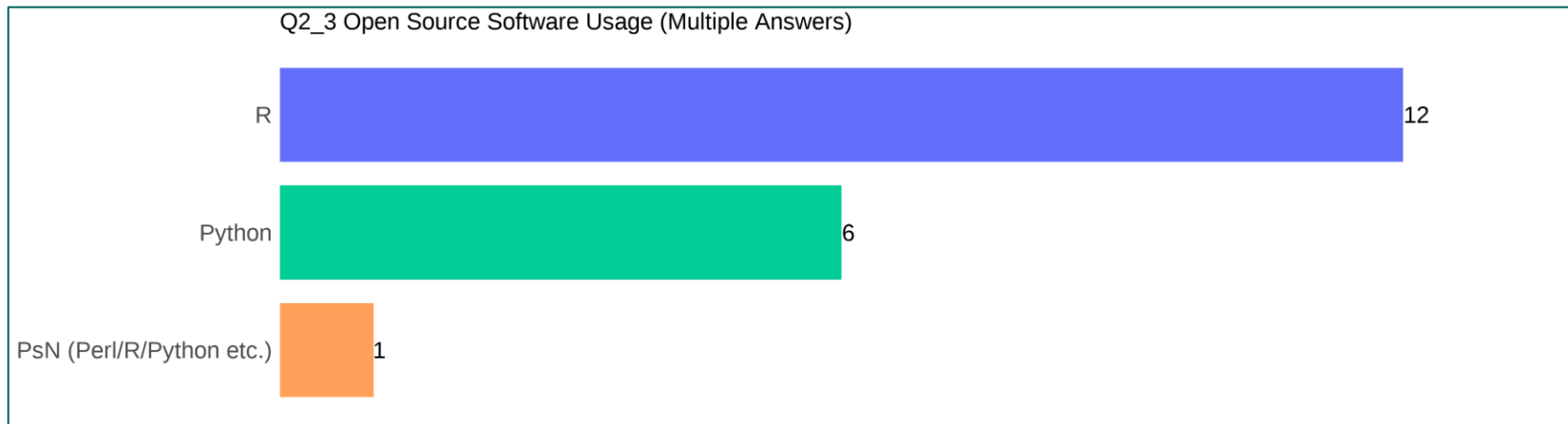
Please tell us about the statistical analysis software you currently use. (Single choice)





Survey Q 2.3

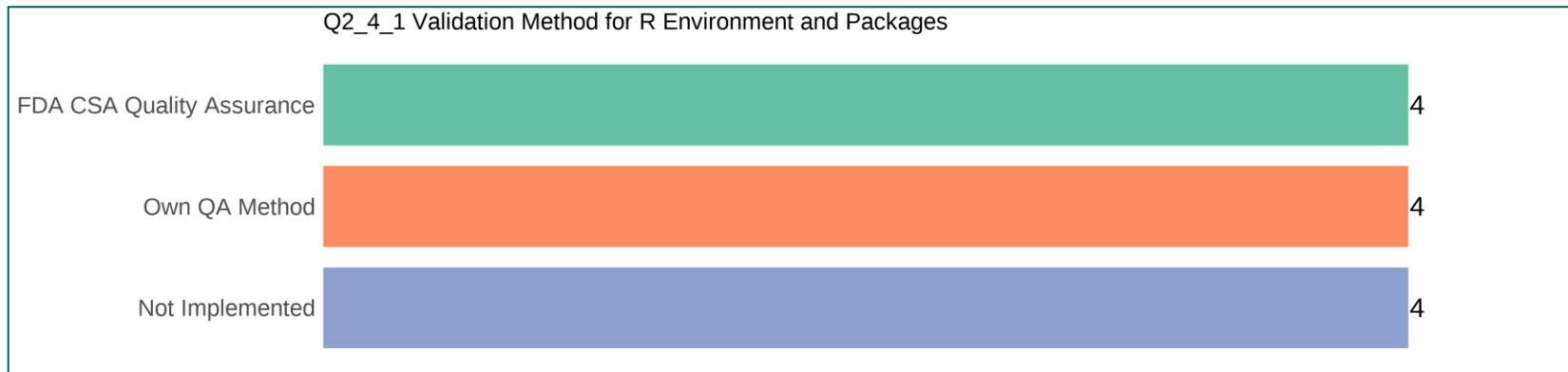
Those who answered that they currently use [OSS] in question 2.1 (including those using it alone or in combination). Please tell us about the OSS you use. (Select all that apply)





Survey Q 2.4.1

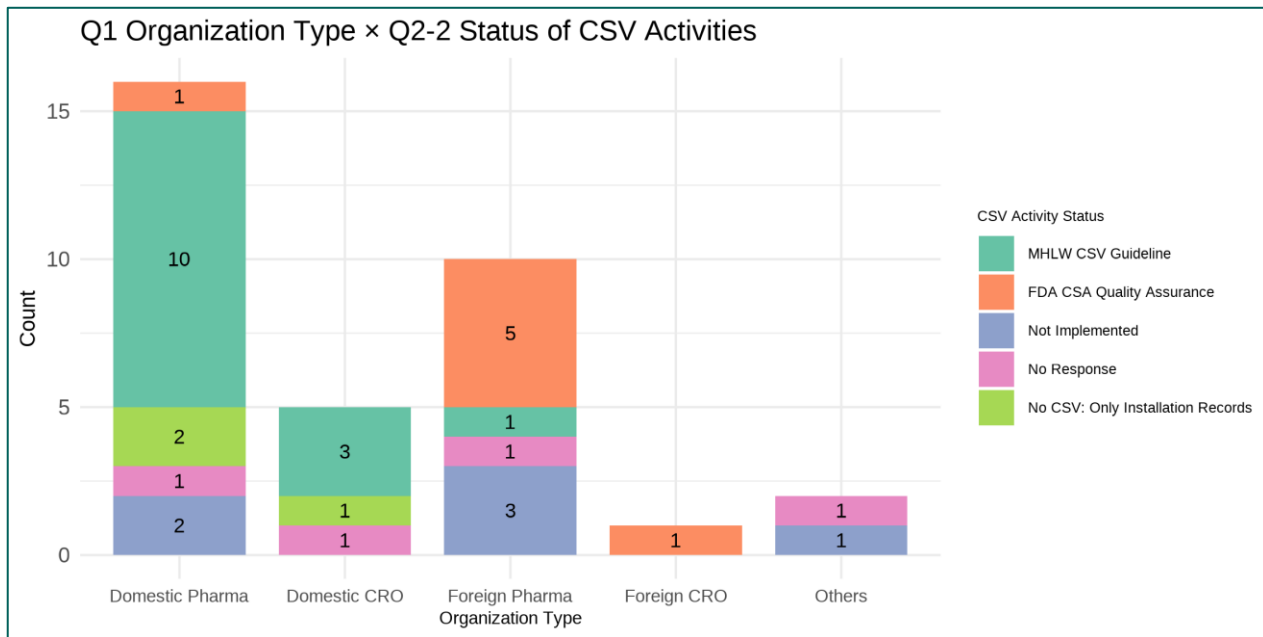
Those who answered that they currently use [R] in question 2.3.
Please tell us about the validation method of R analysis environment and package (Single choice).





Survey Q 2.4.1

By organization type, domestic follows the MHLW CSV guidance and foreign follows the FDA CSA Quality Assurance.



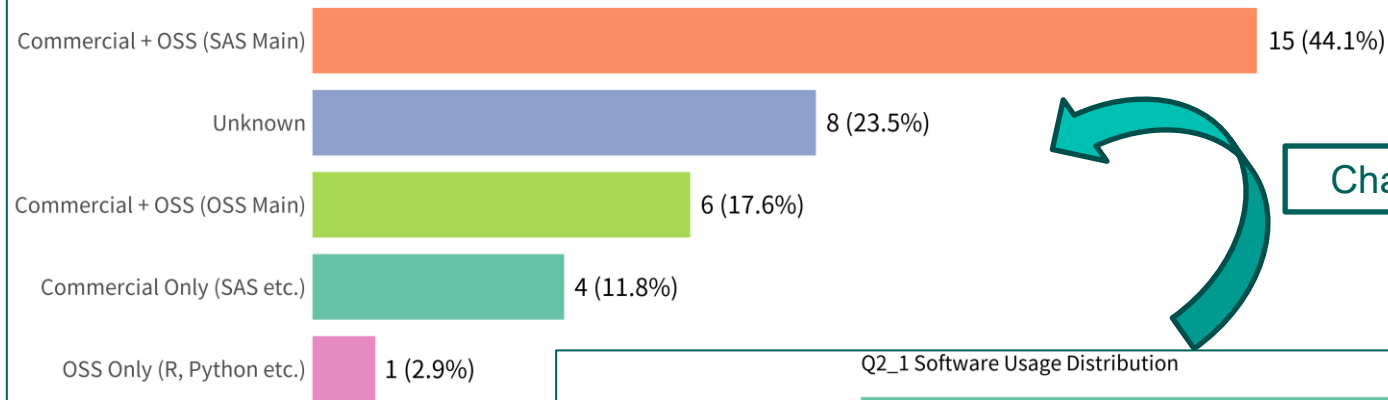
Q3 Future use of [OSS]



Survey Q 3.1

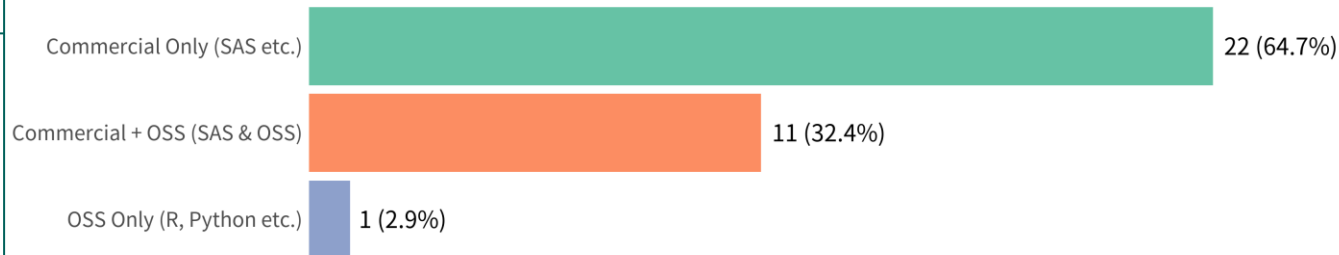
Please tell us about the statistical analysis software you anticipate in 1 to 2 years. (single choice)

Q3 Future Plans for OSS Usage



Changes at 1-2 years

Q2_1 Software Usage Distribution

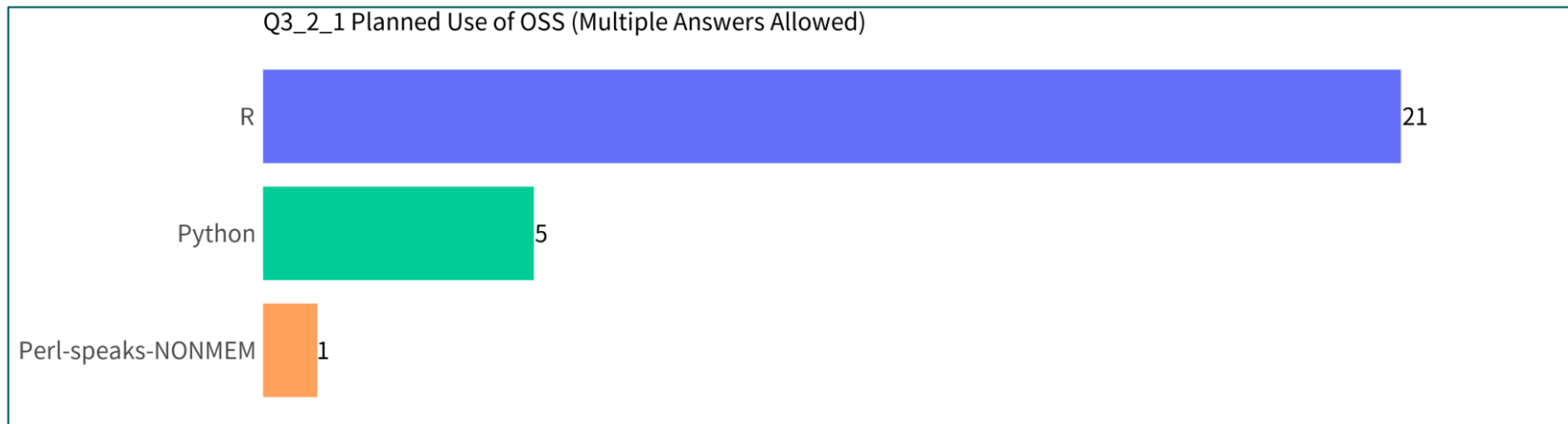




Survey Q 3.2.1

Those who answered that they currently use [OSS] in question 3.1 (including those using it alone or in combination).

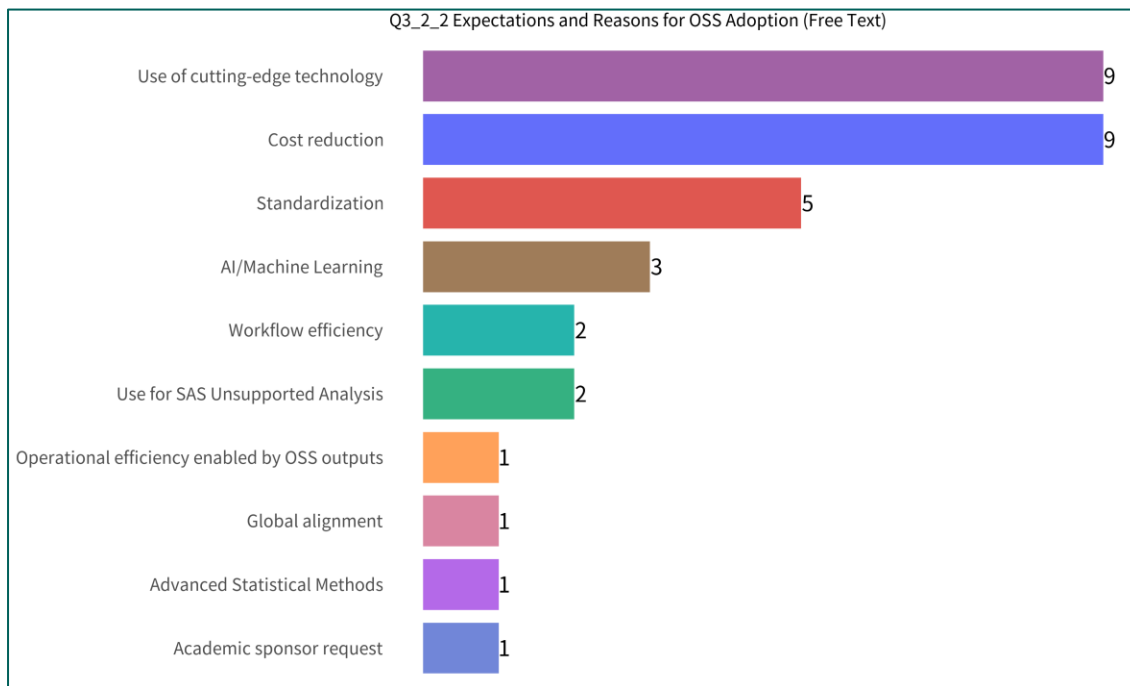
Please provide information on the OSS to be used. (Select all that apply)





Survey Q 3.2.2

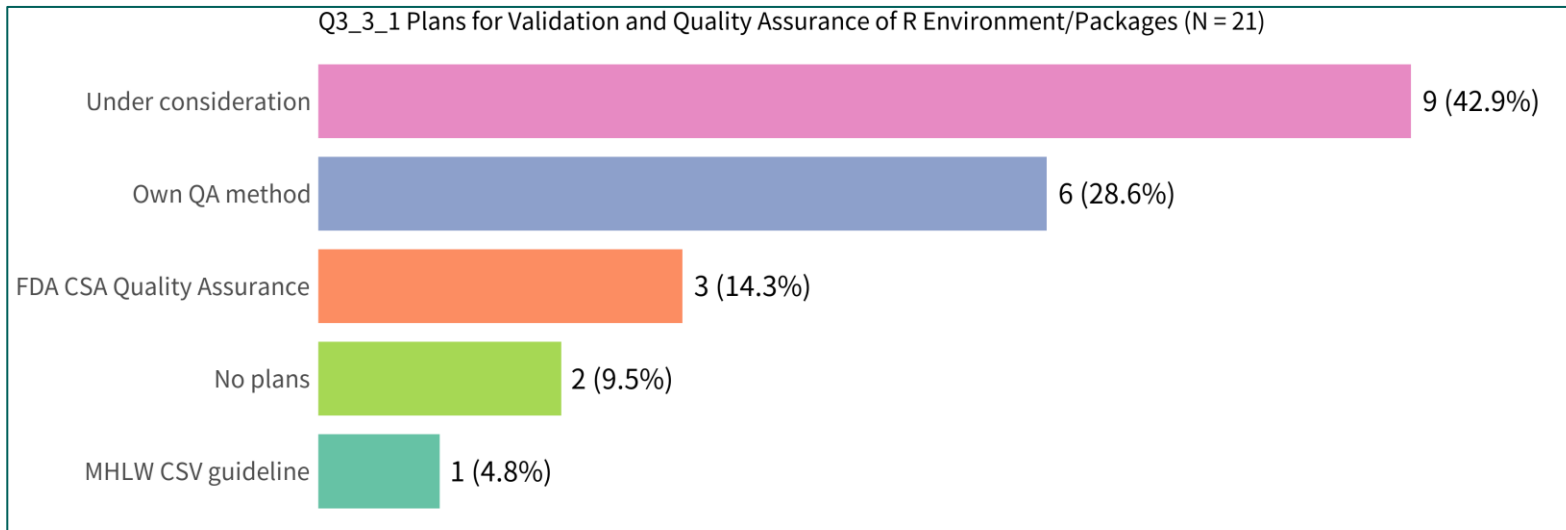
Those who answered that they currently use [OSS] in question 3.1 (including those using it alone or in combination). Please tell us what you expect for OSS and the reason for its introduction. (Free text)





Survey Q 3.3.1

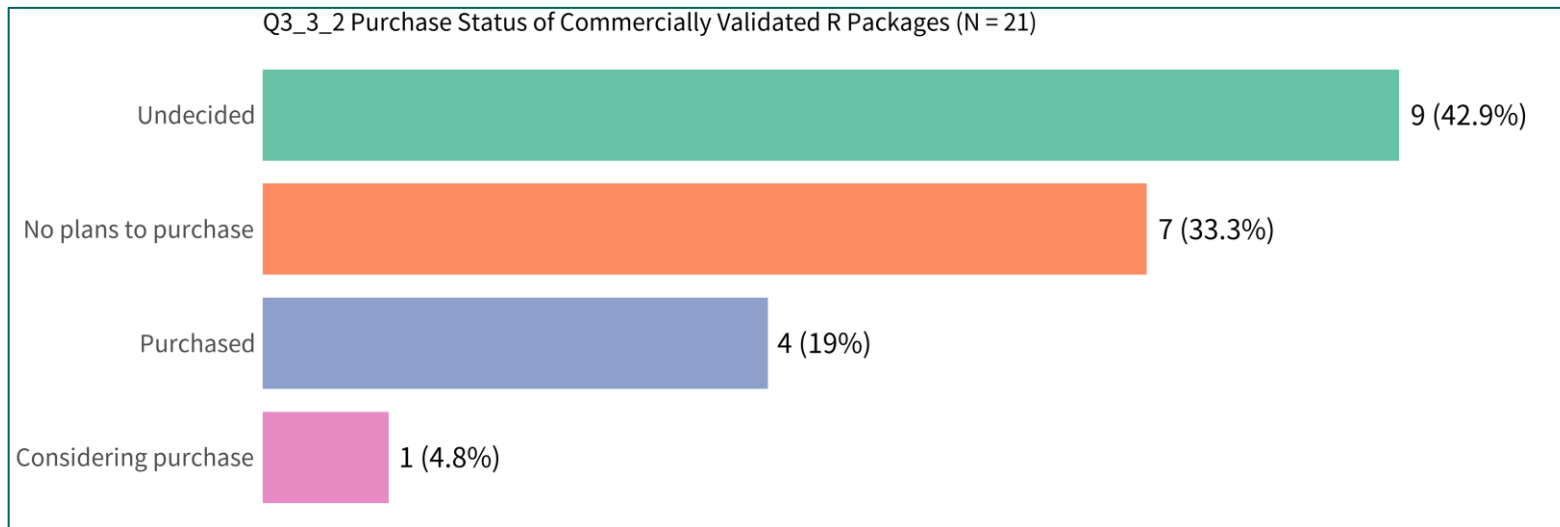
Only for those who plan to use [R] in the future (including those who have already started using R). Please tell us about the plans for validation and quality assurance activities for the R analysis environment and package. (Single choice)





Survey Q 3.3.2

Only for those who plan to use [R] in the future (including those who have already started using R). Please tell us about the purchase status of R packages for which quality control was performed by external vendors with payment. (single choice)



Q4 Quality assurance activities



Survey Q 4

Those who answered [Quality assurance activities have been performed (plan to be performed) by their organization's unique methods] in Q2.4.1 or Q3.3.1,
Please describe the specific method. (free text)

Establishment of evaluation process

Established R package evaluation process
based on R Validation Hub, etc.
Implementation of quality assurance activities
based on R & R packages regarding SOP

Introduction of OSS selection and procurement support system

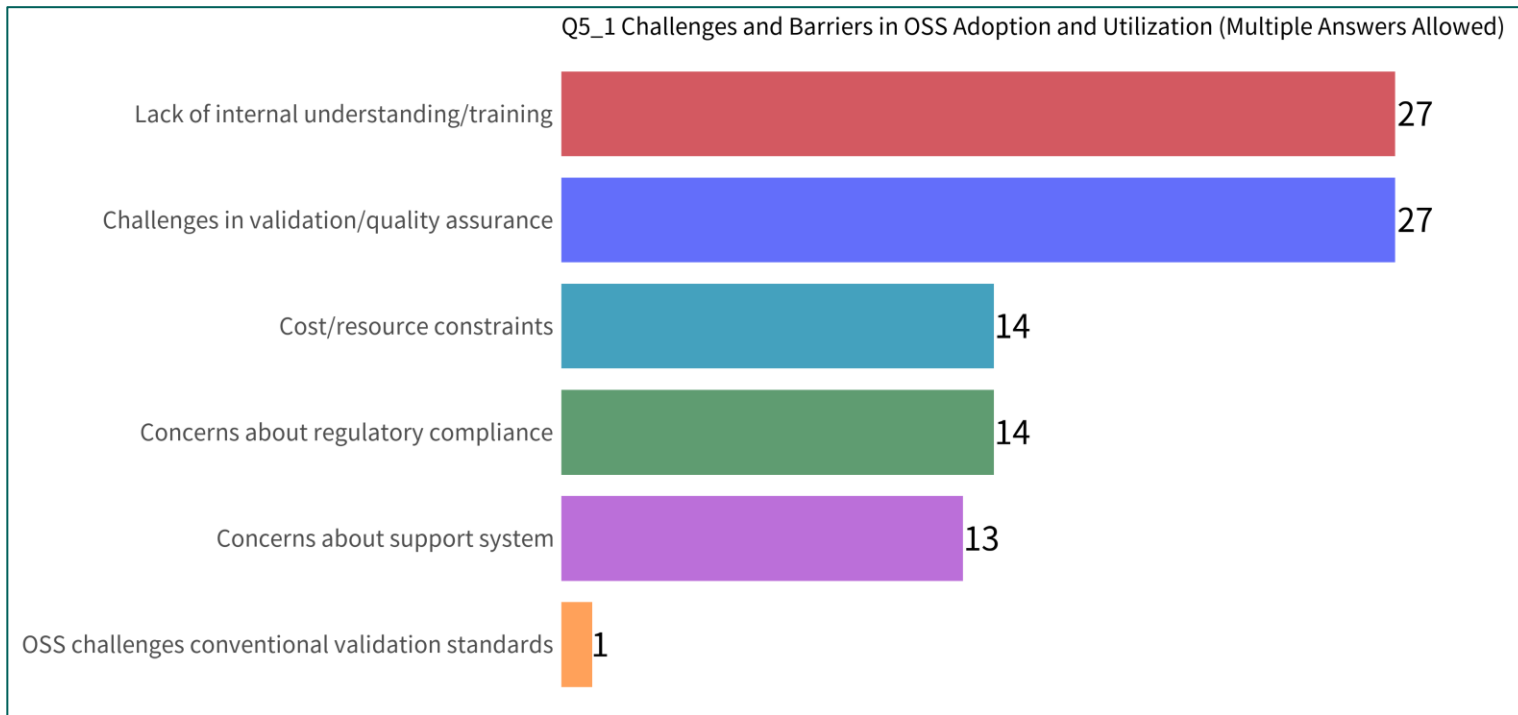
Selection of OSS by own organization /
Implementation of Procurement Support System
(Reliability test of the package has been performed.)

Q5 Issues and barriers in introduction and utilization of OSS



Survey Q 5.1

Please tell us about the issues you feel in introducing and utilizing OSS. (Select all that apply)





Survey Q 5.2

What are some specific examples of issues? Also, please tell us if there are any examples (including specific examples already solved) being considered as solutions to the issues. (free text)

Regulations/quality assurance:

It is difficult to control the version of OSS and ensure the quality assurance of the package, validation, and audit trail, and there are concerns about inconsistency with SAS results and explanation/approval to the authority (PMDA, etc.).

Insufficient education/resources:

There are few people with experience of using R and OSS, and education, training, and sharing of expertise are necessary. Lack of understanding of SOP is also an issue.

Technical issues: There are many technical issues such as rounded specification of R, compatibility of packages, dependency, security risk response, and environmental reproducibility.

Hurdles for introduction:

It is difficult to use existing commercial software, obtain global conformity, and obtain internal approval.

Search for solutions:

We are considering actions such as strengthening of education system, improvement of guidelines and manuals, accumulation of pilot cases, reference to external cases, and utilization of AWS and GIN.

In the future, we aim to solve the issues through education and enlightenment activities, standardization, establishment of a validation system, and consultation with the authorities.

*Summary of free text

R Quality Assurance Opinion Exchange Meeting



R Quality Assurance Opinion Exchange Meeting

- Target: Experts using OSS and R Quality Assurance Team (9 people in total)
- Experts on the use of OSS
 - Those who responded to "Survey on the Status of Use of Statistical Analysis Software and Quality Assurance" and were considered to have abundant experience of using OSS
- Timing: Early October 2025
- Topics for discussion (partial excerpts)
 - Implementation status of quality assurance process for OSS at each company
 - Issues in using OSS
 - ❖ Education and human resources
 - ❖ Outsourcing of Validation
 - ❖ Cost for using OSS



Implementation status of quality assurance process for OSS at each company

- Status of quality assurance of OSS at each company
 - R language and R package are considered individually as quality assurance targets
 - Function that conduct validation differ among companies (e.g. IT, data science, outsourcing)
- R language
 - Treated as identical to other software installed in the statistical analysis environment
- R package
 - Each company has a different policy for quality assurance (in-house or outsourcing)
 - Classified and evaluated according to risk. Those evaluated as low risk can be used in the CSR, while those evaluated as high risk can NOT be used for analysis under GxP.
 - Re-evaluation is required when the package version is updated, and the risk classification may change depending on the results.



Education and human resources

- Education and human resources
 - R education for SAS users is costly, which is one of the hurdles for introduction of OSS.
 - Since new employees have experience of using OSS such as R and python, we expect its spread as the users increase.
- Leveraging industry community
 - Cost reduction and technology sharing are progressing through the use of packages already confirmed by the industry community.
By using R, the industry community can utilize the package whose operation has been confirmed, resulting in a certain level of cost reduction (similar activities are also conducted for SAS).



Outsourcing of validation of R package

- Responsibilities related to validation
 - The responsibility for validation and the risk management system are described in the contract, and the outsourcing expenses are set according to the contents. It may be outsourced, including ongoing software maintenance (e.g., updating R packages).
- Guidelines for performing validation
 - It is unclear which guidelines/guidance, such as the CSV guidelines of the MHLW and the CSA guidance of the FDA, are being followed. Criteria based on the validation report may be included, but prior confirmation as the sponsor is required.

Cost for using OSS

- Advantages and disadvantages of using OSS
 - One of the advantages is that OSS can be used free of charge.
 - However, validation and periodic maintenance are required, and costs (consumption of internal resources) arise to implement them.
 - In addition, if the validation is outsourced, the cost increases further.
 - After examining the use of OSS, we came to believe that the license fee for commercial software includes the cost of validation, etc.

- Summary and future prospects

Summary

- Current status of utilization of OSS
 - While there are many organizations interested in the use of OSS such as R and Python in the industry, there are many issues
 - Key Challenges
 - ❖ Validation methods and guidelines to be followed (including preparation of SOPs, etc.)
 - ❖ Securing resources for validation and periodic maintenance
 - ❖ Technical challenges (compatibility of packages, dependencies, reproducibility, etc.)
- Cost for using OSS
 - It has become clear that validation and periodic maintenance of R language and R package require a certain amount of cost.
Some organizations handle it internally and some organizations outsource it.

Future prospects

- Purpose of using OSS and reconfirmation
 - Use of OSS is a means and not a purpose. Now that it is clear that the use of OSS also has a certain cost, it is necessary to reconfirm whether the organization uses commercially available software or uses OSS as a means to achieve a purpose.
- Education
 - In order for SAS experienced persons to learn R, it is considered to be more efficient to prepare a curriculum for systematic learning than self-learning, but opportunities to learn OSS systematically are limited compared to SAS.
- Promoting industry standards instead of internal standards
 - There is a tendency to promote standards for the entire industry rather than standards unique to each company (similar activities are beginning not only for OSS but also for SAS). Also for validation, since the method and guidelines to follow are not clear, each company should aim to establish the standard process for the industry by sharing examples.



Acknowledgements

- A special thank you to everyone who supported the development of this presentation
 - Masataka Sano (Takushoku University)
 - Hisae Watanabe (ONO PHARMACEUTICAL CO., LTD.)
 - CDISC Japan User Group ADaM members
 - Japan Programming Head Council members
 - Participants of R Quality Assurance Opinion Exchange Meeting

Appendix



Survey Items

- What is your organization?

*Please answer only to this question, including analyses other than those for the clinical study report and CTD

1. Is OSS currently used internally? If yes, tell us the main purpose of use (Select all that apply)

*After this question, respond to the analysis for the clinical study report and analysis for the CTD

2. Current statistical analysis operations (analysis for the clinical study report and analysis for the CTD)

2.1 Please tell us about the statistical analysis software you currently use. (single choice)

2.2 Those who answered that they currently use [commercially available software] in question 2.1 (including monotherapy and combination therapy)

2.2.1 Please tell us about the environmental validation method for commercially available software (SAS, etc.). (single choice)

2.2.2 Please provide information on how to verify analysis results created with commercially available software. (single choice)



Survey Items

2.3 Those who answered that they currently use [OSS] in question 2.1 (including those using it alone or in combination)

Please tell us about the OSS you use. (Select all that apply)

2.4 Those who answered that they currently use [R] in question 2.3

2.4.1 Please tell us about the validation method of R analysis environment and package. (single choice)

2.4.2 Please tell us how to verify the analysis results prepared in R. (single choice)

3. Future use of [OSS]

3.1 Please tell us about the statistical analysis software you anticipate in 1 to 2 years. (single choice)

3.2 Those who answered that they currently use [OSS] in question 3.1 (including those using it alone or in combination)

3.2.1 Please provide information on the OSS to be used. (Select all that apply)

3.2.2 Please tell us expectations for OSS and reasons for its introduction. (free text)



Survey Items

3.3 Only for those who plan to use [R] in the future (including those who have already started using R)

3.3.1 Please tell us about the plans for validation and quality assurance activities for the R analysis environment and package. (single choice)

3.3.2 Please tell us about the purchase status of R packages for which quality control was performed by external vendors with payment. (single choice)

3.3.3 Please tell us who will implement (plan to implement) the validation of R analysis environment/package and quality assurance activities. (Select all that apply)

If you outsource validation or quality assurance activities, please select outsourcing (external vendor or consultant) from the following options.

4. Those who answered [Quality assurance activities have been performed (plan to be performed) by their organization's unique methods] in Q2.4.1 or Q3.3.1.

Please describe the specific method. (free text)



Survey Items

5. Issues and barriers in introduction and utilization of OSS

5.1 Please tell us about issues you feel in introducing/utilizing OSS. (Select all that apply)

5.2 What are some specific examples of issues?

Also, please tell us if there are any examples (including specific examples already solved) being considered as solutions to the issues. (free text)

6. If you have any other opinions or questions, please describe them. (free text)



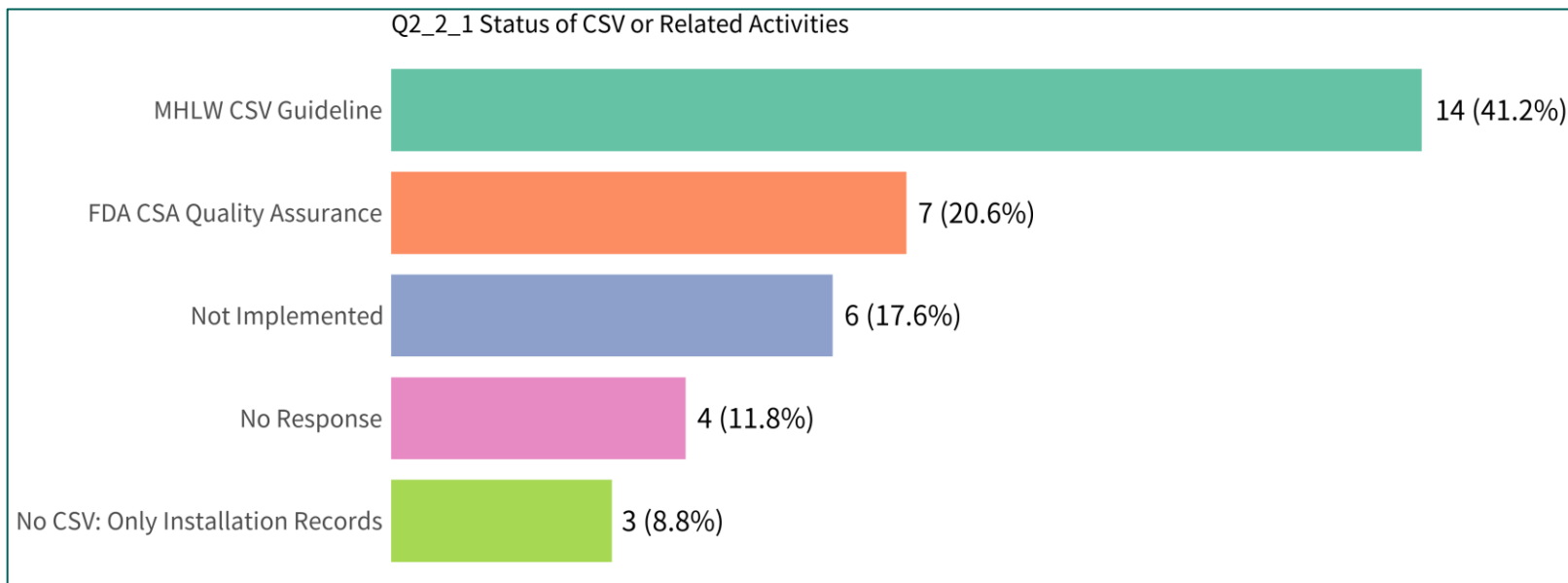
Survey results

- The results of the Survey that could not be posted before the appendix will be displayed here.



Survey Q 2.2.1

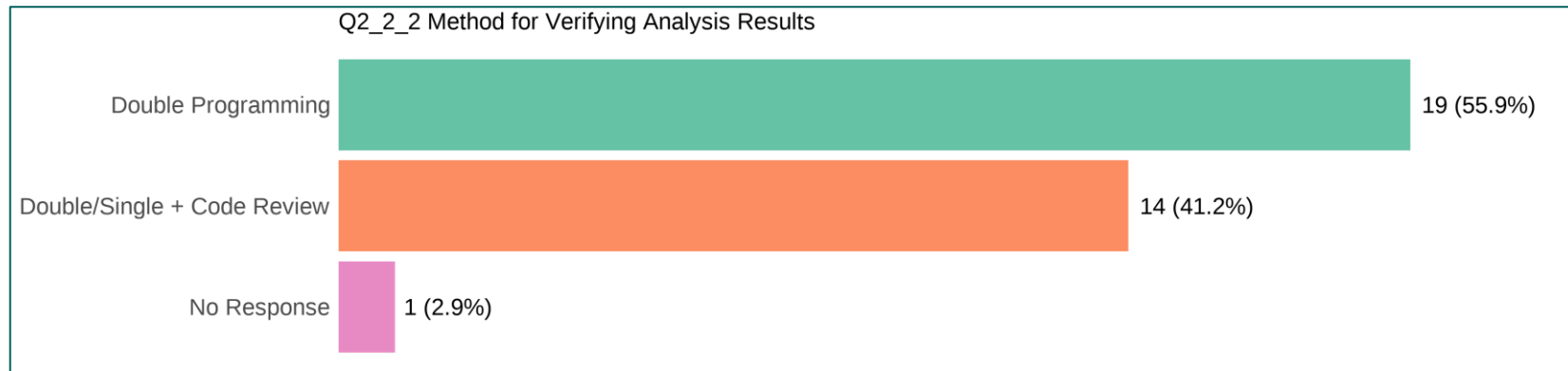
Please tell us about the environmental validation method for commercially available software (SAS, etc.). (Single choice)





Survey Q 2.2.2

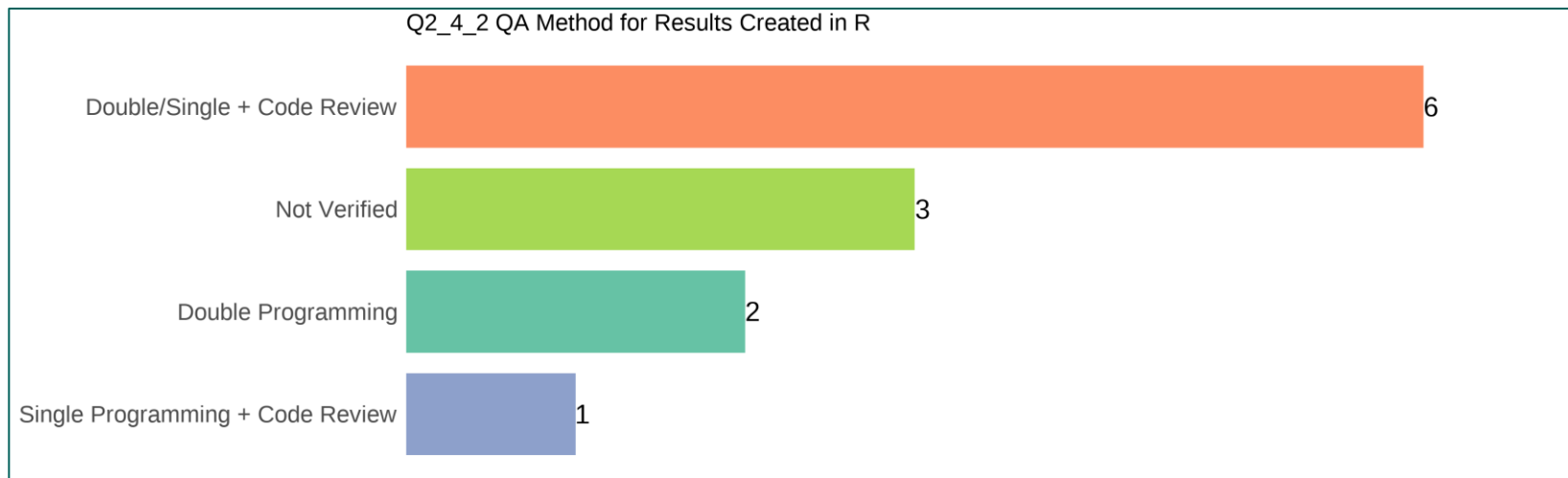
Please provide information on how to verify analysis results created with commercially available software. (Single choice)





Survey Q 2.4.2

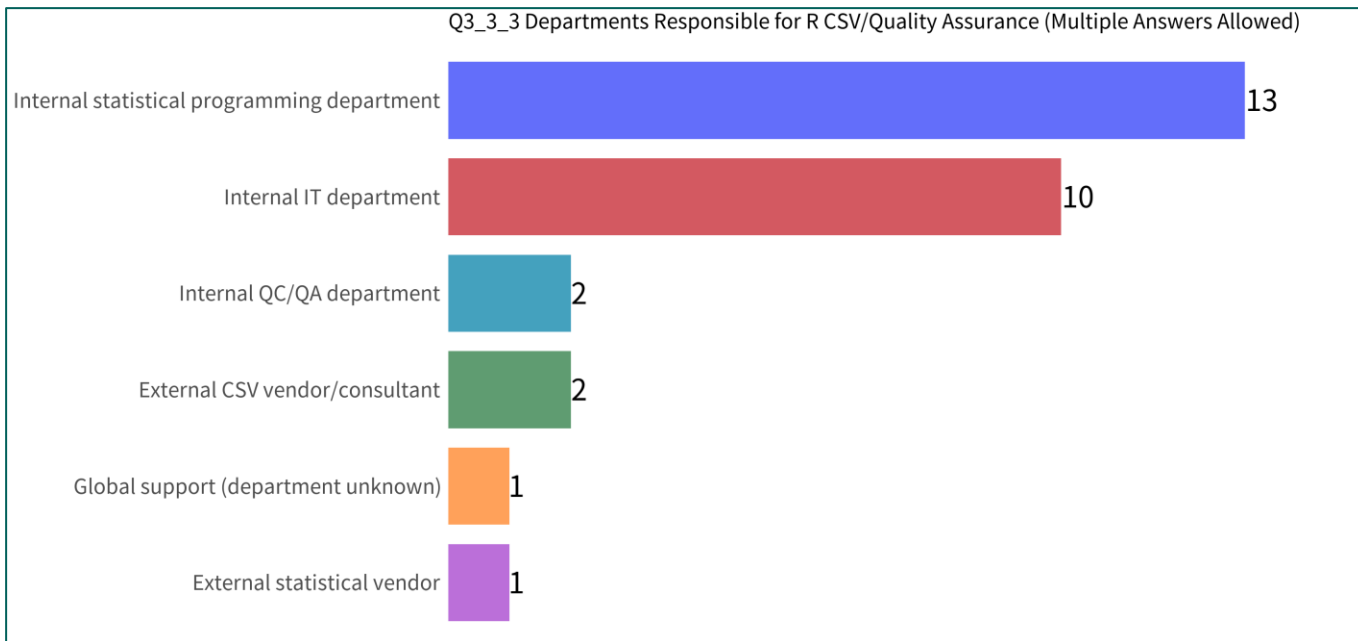
Please tell us how to verify the analysis results prepared in R. (Single choice)





Survey Q 3.3.3

Please tell us who will implement (plan to implement) the validation of R analysis environment/package and quality assurance activities. (Select all that apply) If you outsource validation or quality assurance activities, please select outsourcing (external vendor or consultant) from the following options.





Survey Q 4

Those who answered [Quality assurance activities have been performed (plan to be performed) by their organization's unique methods] in Q2.4.1 or Q3.3.1,
Please describe the specific method. (free text)

- ✓ Based on R Validation Hub, etc., a process for evaluating internal and external R packages is established, and tests, documents, and risk assessments of packages are performed based on the process.
- ✓ OSS Selection and Procurement Support System is available in our organization, and a reliability test of the package was performed.
- ✓ Specific method cannot be provided
- ✓ Risk evaluation and mitigation of the R package will be performed according to the original criteria established by our company.
- ✓ For R and R packages, quality assurance activities are performed in accordance with SOPs. R prepares Validation plan, URS, etc., performs evaluation according to them, and summarizes the results in Validation report. For R packages, a tool to evaluate the risk of the package has been developed and the evaluation results by the tool have been stored (IQ, OQ, PQ, etc. may be implemented for high risk packages). ”



Survey Q 5.2

What are some specific examples of issues? Also, please tell us if there are any examples (including specific examples already solved) being considered as solutions to the issues. (free text)

- ✓ There are commercial R studio (Posit) and Spotfire (TERR tool), but they only guarantee to return the same results for the same input as R, and the contents of the package are not guaranteed.
- ✓ Depending on the R package of the new analysis method, it becomes necessary to upgrade R of the execution environment and compile other packages associated with it, making CSV complicated.
- ✓ PMDA is concerned about whether the quality assurance activities scheduled to be implemented by our company will be sufficiently explainable in the investigation for future application for approval to PMDA and whether the activities will be accepted by PMDA.
- ✓ Since commercially available software is already used, it is difficult to obtain approval from the manager for introduction of OSS from the aspect of validation/quality assurance.



Survey Q 5.2 (cont' d)

What are some specific examples of issues? Also, please tell us if there are any examples (including specific examples already solved) being considered as solutions to the issues. (free text)

- ✓ The training and support system have been improved by holding sessions by a specialized team in the company and experts in functions, enhancing training materials, and frequently updating related information.
- ✓ Since we need to maintain consistency globally, it will take time to consider and introduce it.
- ✓ As the verification of analysis results by OSS, the output is compared with SAS, but it may not always match.
- ✓ Rounding is not standard for rounding in R. Particularly, there are problems in the display of percentages, etc.
- ✓ Many of OSS such as R package are not in version 1.00, and lower compatibility between versions is not guaranteed. In addition, it may be changed to nlmixr2 due to a problem of packages with dependency like nlmixr.
- ✓ If R is used for analysis for CTD, introduction and derivation may be more difficult than in the case where SAS is used (e.g., reproduction of environment, explanation to regulatory authorities).



The Global Healthcare Data Science Community

Contact Channels

UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse