

# Single Day Event

---

## Multilingual Programming: A New Frontier in Clinical Data Analysis

Gloria D'Souza & Yuliana Yakouskaya  
UCB Biosciences



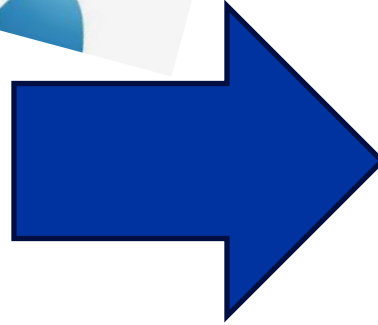
# Disclaimer

The views and opinions expressed in this presentation are solely those of the authors and do not necessarily represent those of UCB

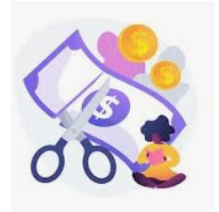


# Need for Multilingual Programming

- Versatility and Flexibility
- Innovation
- Efficiency
- Robust Solutions
- Data Mngmt & Analyses



Medicines out faster,  
efficiently & with  
minimal costs for  
patients





# Advantages of Different Languages

## **SAS**

- Existing expertise and infrastructure
- Verified and validated
- Structured help available
- Customized support
- Optimized for large-scale data processing

## **R**

- Open-source
- Widely used in academic circles
- Most graduates are familiar with R
- Easy to learn
- Some aspects are simpler

## **Python**

- General-purpose programming language
- Open-source
- Web integration
- Easy scripting and automation
- Machine learning
- Large community and formal support

# Challenges & Solutions

## Challenge 1: Transition of SAS programmers to other languages

- Bigger picture
- Start Slow
- Just-in-Time Tailored Training
- Time to practice
- Documentation
- Mentors
- Tandem Language partners
- Hybrid approach



## Challenge 2: Training of new programmers in clinical environment

- Senior-Junior Pairing
- Gradual training



# Challenges & Solutions

## Challenge 3: Transitioning existing projects

WHAT?

WHEN?

HOW?



## Challenge 4: Contingency planning

- Efficient training plans/mentoring
- Future resourcing
- AI implementation



# Practical solutions/applications at UCB

- AI Tool: Programmer Mike
- Patient Profiles
- Python Apps:
  - Define Package
  - ADRG
  - PDF Bundling



Programming Task

Conversion

Source Language

SAS

Target Language

R

## Input

```
proc sort data=adsl;  
  by usubjid;  
run;
```

## Output

```
library(dplyr)  
adsl <- adsl %>% arrange(usrbjid)
```

The `proc sort` procedure in SAS is used to sort a dataset by specified variables. In R, the `dplyr` package provides similar functionality with the `arrange` function. Here, the `adsl` dataset is sorted by the `usrbjid` variable.



# References

[Home - R Consortium \(r-consortium.org\)](https://r-consortium.org)

[Introduction – R Submissions Working Group \(rconsortium.github.io\)](https://rconsortium.github.io)

[How Can Python Be Used in Clinical Trials? \(lexjansen.com\)](https://lexjansen.com)

[SAS: Data and AI Solutions | SAS UK](#)



## The Global Healthcare Data Science Community

### Contact Channels

📞 UK +44 1843 609600  
✉ [office@phuse.global](mailto:office@phuse.global)  
🌐 [phuse.global](http://phuse.global)

### Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)  
📘 [/phusebook](https://facebook.com/phusebook)  
▶ [/phusetube](https://youtube.com/phusetube)  
🌐 [/company/phuse](https://company.phuse.com)