Supplementary Material: *Implementing Code Review in the Scientific Workflow: Insights from Ecology and Evolutionary Biology*

Edward R. Ivimey-Cook*, Joel L. Pick, Kevin R. Bairos-Novak, Antica Culina, Elliot Gould, Matthew Grainger, Benjamin M. Marshall, David Moreau, Matthieu Paquet, Raphaël Royauté, Alfredo Sánchez-Tójar, Inês Silva, and Saras M. Windecker*

*corresponding author(s)
e.ivimeycook@gmail.com & saras.windecker@gmail.com

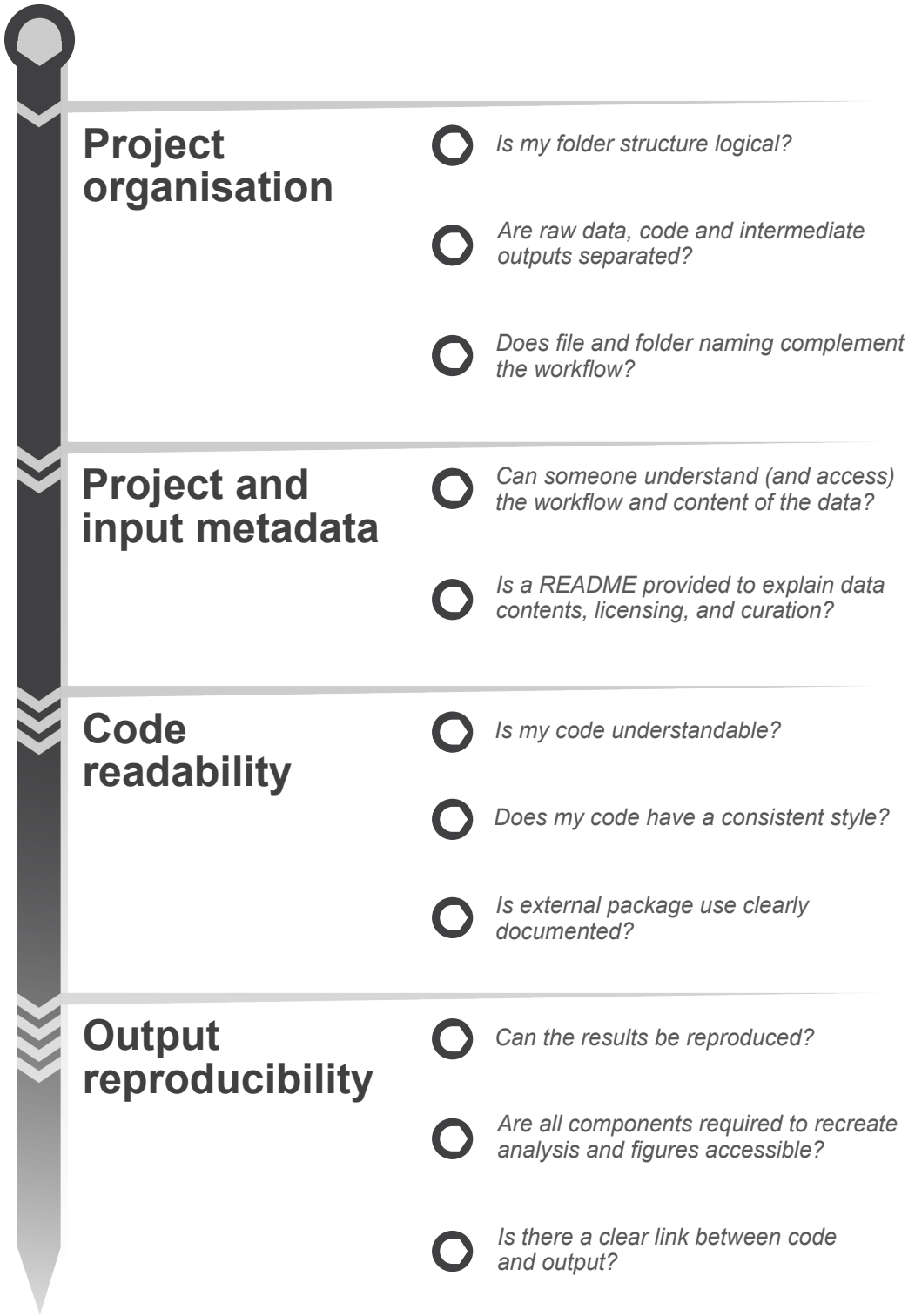

Figure S1. A printable checklist for code review. Figure design by B.M.M.

Box 1: Example 4Rs. *Examples of how each "R" in code review can be assessed.*

1. Is the code as Reported?

For instance, methods may state that an analysis uses a generalised linear model with Poisson error, but the code instead fits a Gaussian error structure.

Methods:
"Data was analysed using a Poisson error structure".

R:
```
lm(response ~ 1 + predictor, family = "Gaussian", data = data)
```

Python:
```
sm.OLS(response, predictor)
```

---

2. Does the code Run?

For instance, code will not be able to be run if it includes calls to libraries (or modules) that are not installed in the current computing environment.

R:
```
library(tidyverse)
Error in library(tidyverse): there is no package called 'tidyverse'.
```

Python:
```
Import numpy as numpy
ModuleNotFoundError: No module named 'numpy'.
```

---

3. Is the code Reliable?

For instance, code may select or modify the wrong column in a dataset, the code will still run, but produce a reproducible yet inaccurate result.

R:
```
wrong.data <- data[,c(1,2)]
right.data <- data[,c("column2", "column3"]
```

Python:
```
wrong.data = data.iloc[:,[0,1]]
right.data = data.loc[:,["column2", "column3"]]
```

---

4. Are the results Reproducible?

For instance, the final outputs may not match those reported in the analysis and results sections (including any relevant figures and narrative text contained within these sections).

Reported result:
"We found a positive effect of *x* on *y*".

Reproduced result:
Reproduced results suggest a negative effect.
Outputs do not match analysis and results.