

Table 6. Distorted Metrics. Here we provide more in-depth information about metric distortion themes identified evaluating tools in Bioconductor (which is ITCR-funded). GEO = Gene Expression Omnibus

Distortion	Example
Accidental Usage	Occasionally scripts used on servers may inadvertently download a package repeatedly and rapidly hundreds to thousands of times, resulting in distorted download metrics that are not representative of real usage. Unique IP download information is useful to distinguish between one user downloading many times versus many users a few times. Given privacy concerns, an alternative solution could involve tracking the timing and approximate location of downloads with a threshold for what would be more than expected as maximum real usage, like a group of people following a tutorial.
Background Usage	There is a baseline background level of downloads across all packages in Bioconductor (including those that are no longer supported). Thus if a new package has 250 downloads in the first year this may seem like a successful number, but actually it is similar to background levels.
Technical vs Research usage	It can be difficult to discern if the usage of a package is for scientific research itself or supporting the implementation of other software. While both are arguably valuable, distinguishing between these motivations can help us understand a particular software's impact in a field. For example, the S4Vectors package [Pagès et al., 2022] is an infrastructure package used by many other packages for technical and non-biological reasons and is therefore not often directly downloaded by end-users. This package is also included in automated checks for other Bioconductor packages using GitHub actions. Another example of support implementation is in the context of container image use. Containerization software (like Docker and Singularity) often install software packages for individual environments that could inflate usage metrics statistics [Merkel, noa, b]. For instance, a user who is actively developing a container may re-trigger the build and thus installation of associated software many times over the course of a project.
Usage Persistence	The affy package [L. et al., 2004], was one of the early packages for microarray analysis, a technology that has largely been replaced by newer technologies, which can be seen by the rate of microarray submissions to GEO overtime. However, despite a the field transitioning away from microarray methods [Mantione et al., 2014], the package was downloaded in 2021 at rates that doubled the rates in 2011. The authors speculate that this could be due to people historically requesting that affy be installed on servers and that this is just persisting, or perhaps it is being used for preliminary hypothesis testing using existing micrarray data, or perhaps it is being used because other microarray packages are no longer supported.