Methods

Emails were sent to scientific authors in November 2018. The email addresses contacted were randomly selected from the MEDLINE baseline 2016, which includes publications from 2017. The addresses were extracted from the Affiliation field from each MEDLINE record. We did not extract emails from the Abstract/Contact section, which only appears in some MEDLINE records. The reference date for each record was the publication date (PubDate).

Overall email address statistics correspond to MEDLINE records up to Nov. 19th, 2018. Free email providers were taken from a list of 4,316 domains maintained in https://github.com/willwhite/freemail/blob/master/data/free.txt. Note that MEDLINE includes records with future publication dates, therefore statistics for 2018 include articles to be published in 2019.

The probability of an email address going stale was derived with a linearly time-dependent Bernoulli process in which the probability of *k* emails bouncing out of *n* emails sent is defined by the equation

$$P(k) = \frac{n}{k} p^k (1-p)^{n-k}$$

in which *p* is a linear function of time,

$$p = \alpha(t_0 - t),$$

where t_0 corresponds to November 2018 and α is a linear coefficient.