The need for large sample numbers to demonstrate lifeless Martian environments

The sample numbers required to achieve a 95% confidence, with high precision, that Martian environments are lifeless can exceed many hundreds. This might not be achieved with sample return only and might need human explorers.

Vanda Inácio[1], Miguel de Carvalho[1,2], Olivia Jackson[3], Sean McMahon[3], Charles S. Cockell[3*]

1. School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh, UK

2. Department of Mathematics & CIDMA, Universidade de Aveiro, Portugal

3. UK Centre for Astrobiology, University of Edinburgh, Edinburgh UK

*corresponding author: c.s.cockell@ed.ac.uk; Tel: +44 131 650 2961

The search for life on Mars is a compelling objective in planetary sciences. In order to make robust statements about life on Mars, we must quantify the number of samples required to achieve given levels of statistical confidence and precision. This is especially the case for a lack of life[1]. It is possible that as we test for evidence of life in samples from Mars, we will be confronted by a growing number of negative or highly ambiguous results. The plausibility of this outcome is amply demonstrated by the controversy surrounding the biogenicity of features observed in Martian meteorite ALH84001[2].

Even if we were to obtain a sample from Mars with unambiguous evidence of life, we would then want to determine specifically which environments were inhabited, which were not, and when. We would need to place some precision with a given confidence on life's absence from certain environments in order to establish the biogeographical distribution of life on Mars and the extent to which it had filled available niche space and physical habitat space. For the sake of speculation, life could have evolved in a Noachian period (4.1 to 3.7 Ga ago) pH-neutral lake such as Gale Crater[3], but failed to evolutionarily radiate into the acidic or saline environments that became more prominent in the Hesperian period (3.7 to 2.9 Ga ago)[4,5]. Such a conclusion would require us to determine how many samples must be found to be lifeless in the latter environments to achieve a given precision in our statements about the probability of life.

The statistical challenge is an unusual one because standard statistical methods, such as those used in medicine to quantify the chances of finding a side effect caused by a new drug, for instance, are based on the known existence of the side effect being sought. In the case of extraterrestrial life, to use a medical analogy, it is similar to speculating that a hitherto unobserved side effect of a drug exists and then attempting to quantify how many human individuals must be examined for this effect before we conclude that the proposed side effect does not exist.

How many lifeless samples?

Statistical methods developed to determine the appropriate sample size for estimating a binomial proportion (in our context, the proportion of samples containing life compared to the total number of samples collected) can be applied to assess the suitability of past, current and future missions to estimate whether Mars lacks life with high precision. We base our sample size calculations on the Clopper–Pearson interval[6]. Unlike many other confidence intervals for binomial proportions, it does not rely on approximating the distribution of the estimated probability of life on Mars using a Gaussian distribution. Instead, calculations are based directly on the binomial distribution.

Consider the case where we have collected n samples, none of which show unambiguous evidence for life and might be taken to be lifeless. In this case of only lifeless samples, the 95% one-sided Clopper–Pearson interval to give us a range of possible probabilities of life is given by[7]:

$$\left(0,\ 1-0.05^{\frac{1}{n}}\right). \qquad (1)$$

That is, given n lifeless samples, we are 95% confident that the probability of life is between 0 and $1-0.05^{\frac{1}{n}}$. So, for instance, if 30 lifeless samples were collected, we would be 95% confident that the probability of life is at most 0.095. In turn, if 500 lifeless samples were collected, the 95% confidence interval would be (0, 0.006), a substantially more precise interval. We note that the upper bound of the interval in (1) is very similar to that derived under a Bayesian approach with a uniform prior on the probability of life[8].

This closed form interval provides a way to calculate the smallest number of samples (n) required such that, with 95% confidence, the probability of life on Mars, or at least within the environment from which the samples have been retrieved, will not exceed a value pre-specified by the researcher, say d, assuming all samples are lifeless. The smaller the value of d, the narrower the interval, i.e., the higher the precision of the interval estimate of the probability of life. Observe that from (1), d is set to $1-0.05^{\frac{1}{n}}$ and solving for n leads to:

$$n=\left\lceil\frac{log(0.05)}{log(1-d)}\right\rceil, \qquad (2)$$

where $\lceil log(0.05)\,/\,log(1-d)\rceil$ is the least integer greater than or equal to $log(0.05)\,/\,log(1-d)$ (the latter may not necessarily be an integer number). Note that before samples are collected, one cannot calculate the confidence interval in (1). The primary goal of this sample size determination exercise is to establish the minimum number of samples that must be collected by missions, to estimate the probability of life by a 95% confidence interval, such that the upper bound of the interval is at most d (or, equivalently, that the probability of life is at most d).

To gain some idea of the samples that would be required, consider an example. Scientists have decided that they require a confidence of 95% in their conclusion and that they wish to be able to state that the probability of life on Mars, or in the environment(s) from which samples have been collected, is 0.005 or less before they are willing to cease collection or announce to the public that it is likely that life did not exist on the planet, or at least in the places examined so far. Thus, if $d = 0.005$, the Clopper–Pearson formulation leads to $n = 598$ samples. The greater the precision of the interval estimate, i.e., the smaller d is, the larger the required number of samples will be. As an example, for $d = 0.001$, we obtain $n = 2995$, whereas $d = 0.0005$ requires $n = 5990$ samples.

These numbers might be considered prohibitively large. However, unfortunately, reducing the confidence of our interval does not escape these order of magnitude results. If a 90% confidence level was instead used (replacing in (1) and (2) 0.05 by 0.1), then for $d = 0.005$, we would require $n = 460$, and for $d = 0.0005$ we get $n = 4605$ samples. Consider relaxing our requirements even further in an attempt to reduce sampling demands. Even for a statistically low and unacceptable confidence level of 50% (replacing in (1) and (2) 0.05 by 0.5), we would still require 139 samples for the largest, less precise, probability width considered ($d = 0.005$). To provide a more complete picture, in Fig. 1 we show the number of lifeless samples required to achieve an upper bound of d (varying from 0.0005 to 0.005) for the interval estimate of the probability of life, when considering a confidence level of 95%, 90%, and 50%.

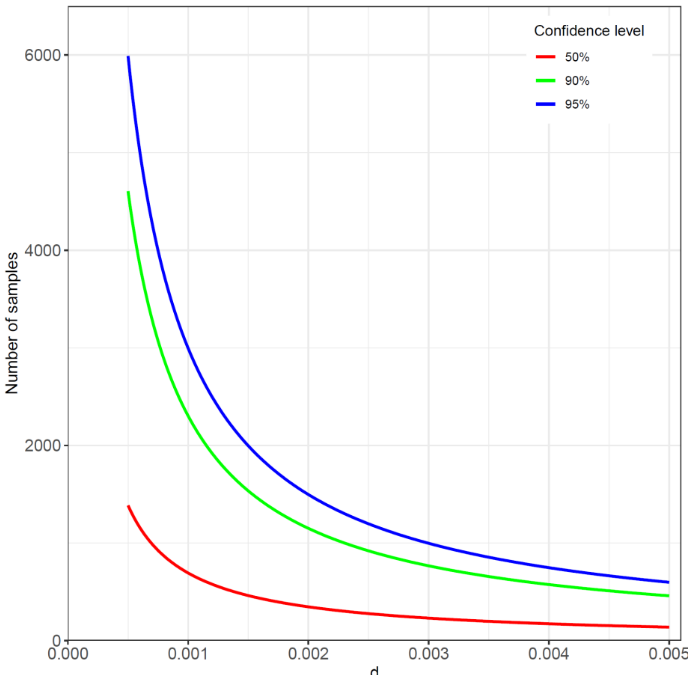

Fig. 1 Determining the lifelessness of Mars and other planetary bodies. Number of lifeless samples needed to achieve an upper bound (*d*) for the interval estimate of the probability of life, for confidence levels of 95% (blue), 90% (green), and 50% (red).

In addition, defining what is a suitable sample to carry out these analyses is not straightforward. For example, centimetre-sized samples (e.g., hand-sized samples) obtained in proximity in the same environment or geological feature might be thought to lack statistical independence, but on the other hand one might argue that such samples contain many subsamples since microorganisms operate at micron-scales. Here, consistent with environmental sampling conventions in microbiology, we consider a sample to be a representative centimetre-scale piece of an environment and we assume that samples need to be collected across the macroscopic region thought to be habitable, which could involve sampling over kilometre scales and in different representative regions.

Furthermore, the total number of samples required will depend on how much information we seek to acquire about the potential for life in distinctive environments, since Mars is a geologically and geochemically heterogeneous world. For example, the numbers we suggest might be collected across all potentially habitable or previously habitable regions across the whole planet, but then they would not provide high statistical precision for particular environments. Instead, these same numbers should be collected for *every* separate region of Mars of a particular environmental type in which life is thought to be possible (e.g., neutral, acidic, saline past aqueous environments[3-5]) to provide high precision for each potential habitat type. Given these factors, the numbers we provide should be considered as a lower bound on the number of samples that may ultimately be needed to achieve the desired precision for the interval estimate of the probability of life in the variety of habitable environments that Mars has hosted and may host today.

Implications for robotic and human exploration

Our results illustrate a qualitatively essential point. If Mars does not yield unambiguous evidence of life, then demonstrating that the planet, or at least the environment(s) studied, lacked life with a precise interval estimate will require a sample number that far exceeds our current robotic mission capabilities. The NASA-ESA Mars Sample return envisages the return of up to 38 samples as well as witness tubes[9]. If we were to assume that its final cache is the potential maximum of 38 samples and that none of them show evidence of life, then the 95%

confidence interval for the probability of life on Mars, based on (1), would be from zero to 0.076. This interval would still leave the question of life significantly unresolved.

One approach would be an expanded sample-return campaign. This might be achieved by international collaboration to launch multiple sample return missions to locations with different geochemical characteristics across the planet. However, the equivalent of many tens of NASA-ESA sample return missions would be needed to achieve the sample numbers we discuss here.

The other approach would be to send human explorers. Even in this case, however, it would not be an easy task: for example, assuming that explorers would collect and analyse three samples a day, and that they wish to achieve a 95% confidence in their conclusion, it will take them around five and a half Earth years to collect and analyse 5990 samples (for $d = 0.0005$), whereas for 598 samples ($d = 0.005$), it will take around six and a half Earth months. In any case, a single dedicated human mission to Mars with long-distance surface transportation capability, allowing for the collection of samples from diverse and promising locations for life across Mars, could potentially achieve high statistical precision of a lack of life.

Although many people consider the discovery of life on Mars to be a potential catalyst for human missions to Mars, an apparent lack of life on Mars may be an even more important driver. The retrieval of a sample with unambiguous evidence for life, for example from sample return, would make it much more likely that limited robotic missions could be used to return further examples of life, removing the need for expensive human missions (although there are other motives for establishing human settlements on Mars[10,11]). However, a growing number of lifeless samples could be a powerful motivator for humans to go to the planet to achieve the statistical confidence sought.

Our approach also shows that current efforts in the development of instrumentation for the search for life on Mars and other planetary bodies, which have focused on improving accuracy and precision, as well as reducing false positive and negative rates[12] cannot overcome the sample number problem, although clearly instruments that are better able to reliably detect biosignatures is a worthwhile objective in itself. Indeed, the presence of false nagatives and positives makes our calculated required numbers of lifeless samples lower bounds. The results we discuss here show not only that instrument design should take into account statistical considerations with respect to performance, but that missions themselves and their sampling capacities should be designed more thoroughly using statistical considerations.

Finally, the analysis we present provides a quantitative way to assess when it would be prudent to relax planetary protection concerns for Mars[13,14] or other planetary bodies, should samples reveal a lack of an extant biota. The decision on what confidence level and probability range would be acceptable to declare a planetary body lifeless from a planetary protection standpoint is a policy decision, but the approach here allows the community to state sample numbers required to achieve certain statistical precision.

Acknowledgements
MdC acknowledges the Royal Society of Edinburgh (Grant 13735861) and FCT (Grants UIDB/04106/2020 / UIDP/04106/2020). CSC acknowledges the Science and Technoilogy Facilities Council (Grant ST/Y001788/1). SM acknowledges the Leverhulme Trust Research Project (Grant RPG-2021-274).

References

1. Cockell, C.S. & McMahon, S. *Nat. Astron.* **3**, 468-470 (2019).
2. McKay, D. S. et al. *Science* **273**, 924–930 (1996).
3. Grotzinger, J.P., et al. *Science*, **343**, 1242777 (2014).
4. Ehlmann, B.L. & Mustard, J.F. *J. Geophys. Res.* **39**, L11202 (2012).
5. Rapin, W. et al. *Nat. Geosci.* **12**, 889-895 (2019).
6. Clopper, C.J. & Pearson, E.S. *Biometrika* **26**, 404–413 (1934).
7. Thulin, M. *Electronic J.Statistics* **8**, 817–840 (2014).

8. Tuyl, F. et al. *The American Statistician* **62**, 40—44 (2008).
9. Meyer, M.A. et al. *Astrobiology* **22**, S5-S26 (2022).
10. Ehlmann, B.L., et al. *Acta Astron.* **56**, 851-858 (2005).
11. Horneck, G. et al. *Adv. Space Res.* **38**, 752-759 (2006).
12. Lorenz, R.D. *Nat. Astron.* **3**, 466-467 (2019).
13. Coustenis A. et al. *Front. Astron. Space Sci.* **10**, 1172546 (2023).
14. Fairén, A.G. & Schulze-Makuch, D. *Nat. Geosci.* **6**, 510-511 (2013).