

# Astro2020 Science White Paper

## Making Exoplanet Surveys Useful for Statistical Population Studies

**Thematic Area:** Planetary Systems

**Principal Author:** Steve Bryson

**Institution:** NASA Ames Research Center

**Email:** [steve.bryson@nasa.gov](mailto:steve.bryson@nasa.gov)

**Phone:** 650-604-2428

**Co-Authors:**

David Bennett, *NASA Goddard and University of Maryland*

Scott Gaudi, *The Ohio State University*

Gijs D. Mulders, *The University of Chicago*

Sharon Xuesong Wang, *Department of Terrestrial Magnetism, Carnegie Institution*

Angie Wolfgang, *The Pennsylvania State University*

**Co-Signers:** Derek Buzasi, *Florida Gulf Coast University*; Douglas A. Caldwell, *SETI Institute*; Joleen K. Carlberg, *STScI*; Jessie L. Christiansen, *Caltech/IPAC*; Courtney Dressing, *University of California, Berkeley*; Eric B. Ford, *Penn State*; Kevin Hardegree-Ullman, *Caltech/IPAC*; Daniel Huber, *University of Hawai`i*; Lisa Kaltenegger, *Carl Sagan Institute Cornell*; Carey Lisse, *Johns Hopkins University Applied Physics Lab*; Benjamin T. Montet, *The University of Chicago*; Timothy Morton, *University of Florida*; Susan E. Mullally, *STScI*; Joshua Pepper, *Lehigh University*; Erik Petigura, *UCLA*; Tae-Soo Pyo, *Subaru Telescope/NAOJ*; Arif Solmaz, *Çağ University*

**Abstract:** In this white paper we discuss what kind of information surveys can provide to facilitate statistical analysis that reveals properties of true exoplanet populations. No single prescription fits all types of surveys. We will gather lessons from various surveys, particularly the *Kepler* mission, which was designed as a survey that results in a statistical occurrence rate, as well as RV and microlensing surveys. We will discuss how diverse and independent surveys can provide data that supports statistical inference. We make several specific recommendations.

## Introduction

The understanding of exoplanet populations is of intrinsic interest (e.g. how common are planets of a particular type?) and provides critical input to both theoretical studies and future mission planning. This understanding arises from a variety of complimentary surveys that provide different windows on exoplanet populations. Radial velocities (RVs) provide exoplanet period and mass, transits provide exoplanet period and size, microlensing provides exoplanet mass, and so on. Different techniques probe different regions of exoplanet properties such as period, mass and size. The combination of these surveys provides the input to statistical studies that provide deep information of the actual distribution of exoplanets and their properties. While no survey probes the entire exoplanet population, combining different surveys and methods provides a comprehensive picture exoplanet demographics, giving insight into the details of planet formation. For example, Clanton & Gaudi (2015, 2016) integrate microlensing, RV, and direct imaging surveys to constrain long-period exoplanets around M-dwarfs and free-floating planets.

No survey is perfect, so survey results cannot be simply assumed to directly or completely describe the actual exoplanet population. However, with proper characterization of a survey, it is possible to correct for imperfections and infer the true exoplanet population. Specifically, surveys are generally *incomplete* (also referred to as *detection efficiency*), missing true exoplanets within the scope of the survey, and are not completely *reliable*, being polluted by false positives masquerading as true planets. They can also be subject to poorly understood *biases*, so that the survey statistics may not match the statistics of the true exoplanet population. Incompleteness, reliability, and bias typically depend on exoplanet properties. The result is that most surveys are biased towards detecting planets with certain ranges of properties. Proper characterization of the incompleteness, reliability and bias of the survey permits a high-confidence statistical inference of the true exoplanet population addressed by the survey.

In this white paper we will discuss what kind of information is required to permit statistical analysis that reveals properties of the true exoplanet population. No single prescription fits all types of surveys. We will gather lessons from various surveys, particularly the *Kepler* mission, which was designed to result in a statistical exoplanet occurrence rate, as well as RV and microlensing surveys. We will discuss how diverse and independent surveys can provide data that supports statistical inference.

Statistical analyses based on previous surveys have revealed unexpected exoplanet population structure, such as the prevalence of sub-Neptunes and nearly coplanar multiple-planet systems, forcing a significant revision of our understanding of planet formation and dynamics. Statistical studies based on future surveys that probe new regions of the exoplanet population can be expected to have similar impacts.

We endorse the findings and recommendations published in the National Academy reports on Exoplanet Science Strategy and Astrobiology Strategy for the Search for Life in the Universe. This white paper extends and complements the material presented therein.

## Lessons From Existing Surveys

### The *Kepler* Mission

The NASA *Kepler* mission was designed as a large-scale survey, with the primary goal of determining the rate of occurrence of Earth-like planets around Sun-like stars (Borucki, et al. 2010). The *Kepler* team and the broader community has provided a series of occurrence rate estimates<sup>1</sup>, revealing population structures such as the prevalence of sub-Neptunes and the evaporation valley, as well as beginning to probe the occurrence of possibly habitable planets. The initial attempts at statistical analysis of the *Kepler* data had significant flaws, such as poor understanding of completeness and reliability. As these occurrence rate estimates were developed, the *Kepler* team evolved a deepening understanding of how to properly characterize the *Kepler* survey's completeness, reliability and biases. Such characterization was complicated by a diverse population of false positives that roughly break into *instrumental false alarms* and *astrophysical false positives*, which meant that optimizing for high reliability severely impacted completeness and vice versa.

As described in Thompson et al. 2018, the *Kepler* team settled on the approach of

- *Uniform automated processing*, so all data was searched and vetted for planet candidates using identical algorithms and data. No data was used that was not available for all searched target stars. Automation avoided bias from human judgement.
- *Characterizing completeness* using modeled transits injected into the observed data, pioneered by Petigura et al. (2013), providing a simulated population of “true” planets for which every ephemeris-matching detection should be vetted as a planet candidate.
- *Characterizing astrophysical false positives* through modeled astrophysical sources.
- *Characterizing instrumental false alarms* by modifying observed data in a way that prevents the detection of true planets and preserves the dominant populations of instrumental false positives.
- *Biases* were primarily addressed by carefully culling the parent target star sample so that the remaining targets in the survey had similar observational coverage. Within the remaining parent sample, relatively small differences in observational coverage was accounted for in the completeness characterization.

The *Kepler* team performed both the survey and completeness and reliability characterization. *Kepler*'s survey data was also characterized by other teams, demonstrating that the group performing the survey may not be the group performing the characterization of completeness, reliability and bias. We therefore abstract out characteristics of the survey data that were required to perform the such characterizations:

- A large subset of the survey's parent population had sufficiently similar observational coverage to be treated as a statistically similar family.
- Low-level as well as highly processed data products were critical to support modeling of astrophysical false positives.
- Observed as well as derived exoplanet properties were published. This allows, for example, the recomputation of planet radii when improved stellar properties became available.

---

<sup>1</sup> [https://exoplanetarchive.ipac.caltech.edu/docs/occurrence\\_rate\\_papers.html](https://exoplanetarchive.ipac.caltech.edu/docs/occurrence_rate_papers.html)

- Non-detections were critical to the statistical analysis.
- Simulated populations of true planets and instrumental false alarms were provided along with an understanding of how well these simulated populations represent their respective actual populations.

This data was sufficient for the *Kepler* team to apply uniform statistical analysis with well-characterized completeness and reliability as well as well-understood biases (though at the time of this writing no peer-reviewed, published occurrence rate making full use of all the *Kepler* completeness and reliability data has appeared). Possibly more importantly, the *Kepler* team provided the completeness and reliability analysis and the underlying data to the community<sup>2</sup>, enabling occurrence rate estimates outside the *Kepler* team.

**The tension between uniformity and accuracy.** One striking feature of the approach used by *Kepler* is the strong imposition of uniformity. This dramatically simplified the characterization of completeness and reliability by treating all elements of the survey population, including both detections and non-detections, in the same way. While there were significant and important follow-up observations that significantly enhanced confidence and understanding of individual planet candidates, because these follow-up observations were not performed on all members of the parent population, using them in the statistical analysis significantly risks introducing unknown biases. For a statistical survey, there should be a predetermined level of observation that is done for all targets and is incorporated into the statistical analysis; in-depth characterization on individual targets to enable individual studies is possible, even encouraged, but should not be incorporated into the statistical analysis.

### Radial Velocity Surveys

Radial velocity surveys roughly fall into two categories: discovery surveys for the purpose of finding previously unknown planets and follow-up surveys that supplement other surveys, for example adding mass measurements to the radius and period measurements from transit surveys. Notable examples of follow-up surveys are the *Kepler* follow-up program and the TESS follow-up program.

Discovery RV surveys typically give significantly more observational coverage to stellar targets with exoplanet detections, and often do not publish non-detections. While appropriate for the goal of exoplanet discovery, this makes these surveys very difficult to integrate into a statistical analysis of exoplanet populations. Lack of knowledge of the parent population makes it very difficult to correct for completeness. Some discovery surveys did provide details on both their selection and observational coverage as well as data about non-detections, e.g. Johnson et al. (2010) and Mayor et al. (2011).

Different issues arise in follow-up RV surveys because their exoplanet population comes from other surveys where completeness and reliability may be known. RV detection significantly enhances the reliability of previously detected exoplanets. Mass measurements, however, may be subject to significant biases:

---

<sup>2</sup> [https://exoplanetarchive.ipac.caltech.edu/docs/Kepler\\_completeness\\_reliability.html](https://exoplanetarchive.ipac.caltech.edu/docs/Kepler_completeness_reliability.html)

- Mass upper limits are rarely published, with the RV community preferring to write papers once a planet’s mass has reached a statistically significant threshold. This leads to a measurable bias in the population’s mass-radius relation (Burt et al. 2018; Montet 2018, Weiss and Marcy 2014), such that very small planets (1-2  $R_{\oplus}$ ) are predicted to be more massive than would be physically expected for that population.
- The decision-making process for which transiting planets are chosen for follow-up is not sufficiently described in the literature. Robust statistical analyses require this process to be reproducible in order to infer accurate population distributions from the observed sample. No published papers describing the mass-radius relation or the small-planet composition distribution take these selection effects into account, partly due to the difficulty in quantifying the time-varying selection function used to create the existing heterogeneous dataset.

### **Microlensing Surveys**

Microlensing surveys have relatively high reliability because, unlike transits and RVs, there are few confounding astrophysical phenomena that mimic a microlensing signal; microlensing false positives are generally excluded through light curve modeling. Further, because the number of detections is a weak function of signal strength one can use high detection thresholds.

Microlensing has long been aware of the need for completeness analysis (Rhie et al. 2000, Gaudi et al. 2002). For example, Suzuki et al. (2016) measured completeness via signal injection, taking care to account for additional observations of actual detections. This survey also considered degeneracies, where the data are consistent with multiple planet models and occasionally binary star models.

### **Generalizations**

The purpose of this white paper is to help survey teams to design their surveys so that they are useful for population studies. We extract the following general lessons:

- Provide details of the parent population, and how the population was selected.
- Complete accounting of observational coverage: Treating the entire parent population with uniform observational coverage, including determination of stellar properties, significantly simplifies the statistical analysis of the survey, but may not always be possible or appropriate. In the cases of non-uniform coverage, details of that coverage such as observation cadence should be published, as well as the reasoning behind the specific choices of a non-uniform coverage (Burt et al. 2018, Suzuki et al. 2016).
- Publish non-detections at the same level of detail as the detections (Burt et al. 2018).
- Provide planet detection tools and descriptions of algorithms such that completeness can be estimated or empirically measured.
- Make both high and low-level data products available for independent analysis, which can be combined with other surveys and ancillary data. For example, Kepler occurrence rates were improved once additional information on target stars became available (e.g., Gaia DR2 release and the California Kepler Survey<sup>3</sup>).

Characterization of completeness and reliability need not be performed by the survey team, so long as appropriate data is provided by the survey. For example, the first use of simulated transit

---

<sup>3</sup> <https://california-planet-search.github.io/cks-website/>

injection in *Kepler* data was performed outside the *Kepler* team (Petigura et al. 2013). Likewise, the TESS mission does not have an explicit goal of statistical population analysis, but is providing sufficient data for the community to perform characterization of completeness, reliability and biases. To assure that it is possible for other teams to characterize completeness and reliability, survey teams that wish to support statistical analysis should assure that they are providing sufficient data for the characterization of completeness and reliability. On the other hand, it is very likely the case that the survey team itself is best positioned to characterize the completeness and reliability of its survey.

We believe that properly characterized surveys will enable significant progress in the understanding of exoplanet population statistics. Once a survey's completeness, reliability and bias are characterized it will be relatively simple to integrate the survey results with other, similarly characterized surveys even if the surveys are based on completely different observational methods.

## **Recommendations and Opportunities**

We make the following recommendations to support the statistical analysis of exoplanet survey results:

- Surveys should be clear and explicit about whether or not they are intended to support statistical analysis, at least in part.
- For surveys intended to support statistical analysis, survey design should include the provision of sufficient data for the characterization of completeness, reliability and bias. Attempting to support such characterization as an afterthought can be significantly more difficult and costly. Survey teams need not perform the characterization itself, but should show that the data provided by the survey is sufficient for other groups to perform that characterization.
- Surveys should clarify their selection criteria for planet detection and follow-up, documenting the depth and homogeneity of their samples.
- Surveys should publish non-detections.
- Funding agencies and reviewers should encourage and reward surveys that provide a plan for supporting statistical analysis.

Surveys designed to support statistical analysis will enable science possibly well beyond the specific goals of the survey. We see several opportunities:

- Surveys support each other. If a survey's results can be translated into properties of the actual exoplanet population via statistical inference, then that survey's results can be directly compared to other similarly characterized surveys. Such a comparison provides consistency checks and constraints on the exoplanet population. Several surveys have pointed out the consistency of their results with other surveys where they overlap.
- Similarly, several well-characterized surveys can be used to perform an integrated statistical analysis using results from multiple, diverse survey types. Such an analysis would provide understanding of exoplanets with a range of properties that cannot be covered by the surveys taken individually.

An example of the kind of science enabled by well characterized surveys is found in the companion white paper *Wide Orbit Exoplanet Demographics* by Bennett et al.

## References

- Burt, J., Holden, B., Wolfgang, A., et al. 2018, *AJ*, **156**, 255.
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977.
- Clanton, C., & Gaudi, B. S. 2016, *ApJ*, **819**, 125.
- Clanton, C., & Gaudi, B. S. 2017, *ApJ*, **834**, 46.
- Gaudi, B. S., Albrow, M. D., An, J., et al. 2002, *ApJ*, **566**, 463.
- Johnson, J. A., Aller, K. M., Howard, A. W., et al. 2010, *PASP*, **122**, 905.
- Mayor, M., Marmier, M., Lovis, C., et al. 2011, arXiv e-prints , arXiv:1109.2497.
- Montet, B. T. 2018, *RNAS*, **2**, 28.
- Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, *PNAS*, **110**, 19273.
- Rhie, S. H, Bennett, D. P., Becker, A. C., et al. 2000, *ApJ*, **533**, 378.
- Suzuki, D., Bennett, D. P., Sumi, T., et al. 2016, *ApJ*, **833**, 145.
- Thompson, S. E., Coughlin, J. L., Hoffman, K., et al. 2018, *ApJS*, **235**, 38.
- Weiss, L. M., & Marcy, G. W. 2014, *ApJ*, **783**, L6.