Draft version February 23, 2018
Typeset using LaTeX **modern** style in AASTeX61

BINARITY'S BIASES ON TRANSIT SURVEY OCCURRENCE RATES

L. G. Bouma,[1] K. Masuda,[1] and J. N. Winn[1]

[1] *Department of Astrophysical Sciences, Princeton University, 4 Ivy Lane, Princeton, NJ 08540, USA*

Submitted to AAS journals.

ABSTRACT

Wide-field surveys for transiting planets, such as the NASA *Kepler* and *TESS* missions, are usually conducted without knowing which stars have binary companions. Unresolved and unrecognized binaries give rise to systematic errors in planet occurrence rates, including misclassified planets and mistakes in completeness corrections. The individual errors can have different signs, making it difficult to anticipate the net effect on inferred occurrence rates. Here we use simplified models of signal-to-noise limited transit surveys to try and clarify the situation. We derive a formula for the apparent occurrence rate density measured by an observer who falsely assumes all stars are single. The formula depends on the binary fraction; the mass function of the secondary stars; and the true occurrence of planets around primaries, secondaries, and single stars. It also takes into account the Malmquist bias by which binaries are over-represented in flux-limited samples. Application of the formula to an idealized *Kepler*-like survey shows that for planets larger than $2\,R_\oplus$, the net systematic error is of order 5%. In particular, unrecognized binaries are unlikely to be the reason for the apparent discrepancies between hot Jupiter occurrence rates measured in different surveys. For smaller planets the size the errors are potentially larger: the occurrence of Earth-sized planets could be overestimated by as much as 50%. We also show that whenever high-resolution imaging reveals a transit host star to be a binary, the planet is usually more likely to orbit the primary star than the secondary star.

Keywords: methods: data analysis — planets and satellites: detection — surveys

Corresponding author: L. Bouma
luke@astro.princeton.edu

1. INTRODUCTION

One of the goals of exoplanetary science is to establish how common, or rare, are planets of various types. Knowledge of planet occurrence rates is helpful for inspiring and testing theories of planet formation, designing the next generation of planet-finding surveys, and simply satisfying our curiosity. One method for measuring occurrence rates is to monitor the brightnesses of many stars over a wide field, seeking evidence for planetary transits. This was the highest priority of the NASA *Kepler* mission. Great strides have been made in the analysis of *Kepler* data, including progress towards measuring the fraction of Sun-like stars that harbor Earth-like planets (Youdin 2011; Petigura et al. 2013; Dong & Zhu 2013; Foreman-Mackey et al. 2014; Burke et al. 2015).

A lingering concern about these studies is that in most cases, investigators have assumed that all of the targets in the survey are single stars (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Dressing & Charbonneau 2015; Burke et al. 2015). In reality, many of the sources that are monitored in a transit survey are unresolved multiple-star systems, mainly binaries. Unrecognized binaries cause numerous systematic errors in the planetary occurrence rates. For example, when there is a transiting planet around a star in a binary, the additive constant light from the second star reduces the fractional loss of light due to the planet. This makes transit signals harder to detect and lowers the number of detections. On the other hand, a binary system presents two opportunities to detect transiting planets, which could increase the overall number of detections.

At the outset of this study it was not clear to us whether the neglect of binaries is a serious problem, or even whether the net effect of the errors is positive or negative. The goals of this study were to provide a framework for dealing with these issues, and to gauge at least the order of magnitude of the systematic errors. In this spirit, our models are idealized. We do not attempt a detailed correction of the results from *Kepler* or any other real transit survey. We took inspiration from the insightful analytic model for transit surveys by Pepper et al. (2003).

This paper is organized as follows. The next section enumerates the various errors that arise from unrecognized binaries. Then in Section 3, we develop an idealized model of a transit survey in which all planets have identical properties, and all stars are identical except that some fraction are in binary systems. This simple model motivates the derivation of a general formula, given in Section 4, that allows for more realistic stellar and planetary populations. We use this formula in Section 5 to explore more realistic models. We discuss the errors due to unrecognized binaries for specific cases of current interest: the occurrence of Earth-like planets; the apparent discrepancy between hot Jupiter occurrence rates measured in different surveys; and the shape of the "evaporation valley" in the planet radius distribution that was brought to light by Fulton et al. (2017). We summarize and discuss all the results in Section 6.

2. UNDERSTANDING THE ERRORS

Imagine that a group of astronomers wants to measure the mean number of planets per star. They are particularly interested in planets of radius r and stars of mass M and radius R. They obtain a time series of images of some region of the sky, and prepare light curves for a large number of unresolved sources. Then they search these light curves for transit signals and detect all the signals for which

$$\frac{\delta}{\sigma} > \left(\frac{\mathrm{S}}{\mathrm{N}}\right)_{\mathrm{min}}. \tag{1}$$

Here the signal, δ, is the observed transit depth, the dimensionless fraction by which the total light fades during transits. Note that although δ is often equated with $(r/R)^2$, this is not true when the host star is a member of an unresolved binary. In those cases, δ is smaller than $(r/R)^2$ because of the constant light from the binary companion, an effect often called the "dilution" of the transit signal. The noise, σ, is the fractional uncertainty in the determination of the flux of the source, which may include multiple stars that are blended together. The threshold signal-to-noise ratio depends on the desired level of confidence that the signal is real.

The astronomers analyze their data assuming that all the sources are single stars. In particular they do not have accurate enough parallaxes to tell that some of the stars appear to be overluminous. They count the number N_{det} of transit signals that appear to be produced by the desired type of planet around the desired type of star. They also count the number $N_\star$ of "searchable stars" in their survey, i.e., the number of stars of the desired type that are bright enough to have allowed for the detection of a transit signal with amplitude $(r/R)^2$. They estimate the occurrence rate to be

$$\Lambda = \frac{N_{\mathrm{det}}}{N_\star}\,\frac{1}{p_{\mathrm{tra}}}, \tag{2}$$

where the geometric transit probability, p_{tra}, accounts for the fact that most planetary orbits are not aligned close enough with our line of sight to produce transits.

There are many potential pitfalls in this calculation. Some genuine transit signals are missed even if they formally exceed the signal-to-noise threshold, because of the probabilistic nature of transit detection. Planets can be misclassified due to statistical and systematic errors in the catalogued properties of the stars. Some transit-like signals are spurious, arising from noise fluctuations or failures of "detrending" the astrophysical or instrumental variations in the photometric signal. Poor angular resolution leads to blends between eclipsing binary stars and other stars along nearly the same line of sight, producing signals that mimic those of transiting planets.

Here, though, we will focus exclusively on problems that arise from the fact that many stars exist in gravitationally bound binary systems. We will also focus on the errors in planet occurrence as a function of radius, rather than orbital period. This is because when more than one transit is detected (as is usually required by

the surveyors), the orbital periods can be measured without ambiguity regardless of whether the host star is single or one member of a binary. Even with this narrow focus, there are numerous sources of error. All three of the quantities in Equation 2 are biased:

1. The number of detected planets, $N_{\rm det}$, is actually the number of detected planets that *appear* to have size r, orbiting stars that *appear* to have mass M. Whenever the planet-hosting star is part of a binary,
    - the planet's size could be misclassified because of the reduction in the amplitude of the photometric signal;
    - the host star's properties could be misclassified because its light is combined with a second star of a different spectral type.
2. The number of searchable stars, $N_\star$, is biased
    - toward lower values, because it does not include all of the secondary stars that were inadvertently searched for transiting planets;
    - toward higher values, because some of the stars that appeared to be searchable are in fact binaries for which the amplitude of the photometric signal would have been reduced to an undetectable level.
3. The transit probability $p_{\rm tra}$ is biased because the planet-hosting star could be misclassified. At fixed orbital period, the transit probability is proportional to $\rho^{-1/3}$, where ρ is the stellar mean density (Winn 2010). Therefore, any errors in determining the host star's mean density lead to errors in the correction for the transit probability.

There are at least two other complications that may arise, which are not represented in Equation 2. The first one is an observational effect. Within the sample of apparently searchable stars, the ratio between the number of binary and single stars will differ from the ratio that would be found in a volume-limited sample. This is due to a type of Malmquist bias. The total luminosity of a binary is larger than the luminosity of either the primary or secondary star. This means that for transit signals of a given amplitude, sources that are binaries appear to be searchable at greater distances from the Earth. Binaries are therefore over-represented in the collection of apparently searchable stars.

The other complication is astrophysical: the true occurrence rate of a certain type of planet may depend on whether the host is a single star, the primary star of a binary, or the secondary star of a binary. The rate might also depend on the characteristics of the binary, such as the mass ratio and orbital period. Such differences could be caused by the requirement for long-term dynamical stability, or differences in the planet formation process. When the search sample includes both singles and binaries,

the detected planets are thereby drawn from different occurrence distributions (see Wang et al. 2015a; Kraus et al. 2016).

Given all of the confusing and opposing sources of error, we will proceed in stages. We start with a model so simple that everything can be written down on the back of a napkin, and build up to an analytic model allowing for generality in the distribution of the binaries and the planets they host.

3. SIMPLE MODELS

3.1. *One type of star, one type of planet*

Since the effects of binarity are most pronounced when the two stellar components are similar, we begin by considering a universe in which all stars are identical, with mass M, radius R, and luminosity L. Single stars are uniformly distributed in space with a number density of $n_{\rm s}$ stars per cubic parsec, and binaries are uniformly distributed with number density $n_{\rm b}$. In this scenario, stars are never misclassified because the combined light of a binary has the same color and spectrum as a single star. We further assume that all planets have the same radius, r, and occur around single stars and members of binaries at the same rate, $\Lambda(r)$.

Our naive observers conduct a transit survey. To calculate the occurrence rate of planets with radius r, they count the number of detections of signals with amplitude $(r/R)^2$. Then they identify all the sources that appear to have been searchable for a signal of amplitude $(r/R)^2$. Here and throughout the rest of this paper, we assume that the limiting source of noise is the photon-counting noise from the source, i.e., $\sigma \propto 1/\sqrt{F}$, where where F is the total flux of the source. Thus the observers determine the minimum F_0 for which detection would have been possible, and count the number of sources with $F_{\rm tot} > F_0$. This will include all the single stars out to a maximum distance

$$d_0 = \sqrt{\frac{L}{4\pi F_0}}. \tag{3}$$

Since binaries are twice as luminous, the condition $F > F_0$ will include binaries out to the larger distance of $d_0\sqrt{2}$. None of the stars in binaries will appear to have a planet of radius r, because of the dilution of the transit signal. Thus, the *apparent* occurrence rate $\Lambda_{\rm a}$ of planets of radius r is

$$\Lambda_{\rm a}(r) = \frac{\Lambda(r)\, n_{\rm s} d_0^3}{n_{\rm s} d_0^3 + n_{\rm b}(d_0\sqrt{2})^3}. \tag{4}$$

This apparent rate is larger than the true rate by a factor

$$\frac{\Lambda_{\rm a}(r)}{\Lambda(r)} = \frac{1}{1 + 2^{3/2}(n_{\rm b}/n_{\rm s})}. \tag{5}$$

The observers will also detect some transiting planets around stars with binary companions. The amplitude of these signals is

$$\frac{L(r/R)^2}{L+L} = \frac{1}{2}\left(\frac{r}{R}\right)^2 = \left(\frac{r/\sqrt{2}}{R}\right)^2, \tag{6}$$

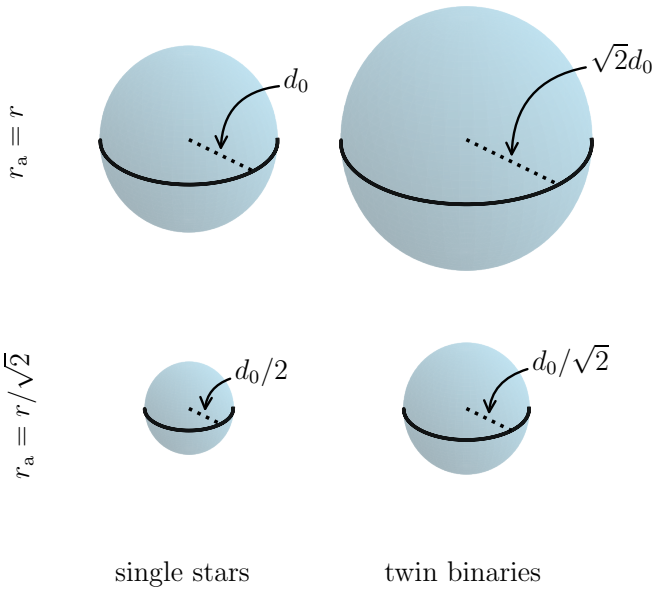

Figure 1. *Top.*—Volumes within which sources appear to be searchable for planets of radius r. Single stars are searchable out to a distance d_0, at which point they become too faint to allow the detection of transits. Binaries are brighter and therefore appear to be searchable out to a larger distance. *Bottom.*—Volumes within which within which stars appear to be searchable for planets of radius $r/\sqrt{2}$.

leading the astronomers to believe they have discovered a population of planets with radius $r/\sqrt{2}$. To calculate the corresponding occurrence rate, they count the stars for which this type of signal would have been detectable. The limiting flux for detection in this case is $4F_0$, because the signal amplitude hase been reduced by a factor of two and the noise level must also be reduced by a factor of two. The condition $F > 4F_0$ is met for single stars within a distance $d_0/2$, and binaries within a distance $d_0\sqrt{2}/2$. Therefore, the observers will calculate the occurrence rate of this new type of planet to be

$$\Lambda_{\rm a}\left(\frac{r}{\sqrt{2}}\right) = \frac{2\Lambda(r)\,n_{\rm b}(d_0\sqrt{2}/2)^3}{n_{\rm s}(d_0/2)^3 + n_{\rm b}(d_0\sqrt{2}/2)^3} = \frac{2\Lambda(r)\cdot 2^{3/2}(n_{\rm b}/n_{\rm s})}{1+2^{3/2}(n_{\rm b}/n_{\rm s})}. \tag{7}$$

Figure 1 illustrates the volumes enclosing the apparently searchable single and binary stars.

We can now assess the severity of the errors, for a given value of the binary-to-single ratio $n_{\rm b}/n_{\rm s}$. For stars with masses from 0.7 to 1.3 $M_\odot$, Raghavan et al. (2010) found the multiplicity fraction – the fraction of systems in a volume-limited sample that are multiple – to be 0.44. Assuming all multiple systems are binaries, this gives a binary fraction

$$\frac{n_{\rm b}}{n_{\rm s}+n_{\rm b}} \approx 0.44, \tag{8}$$

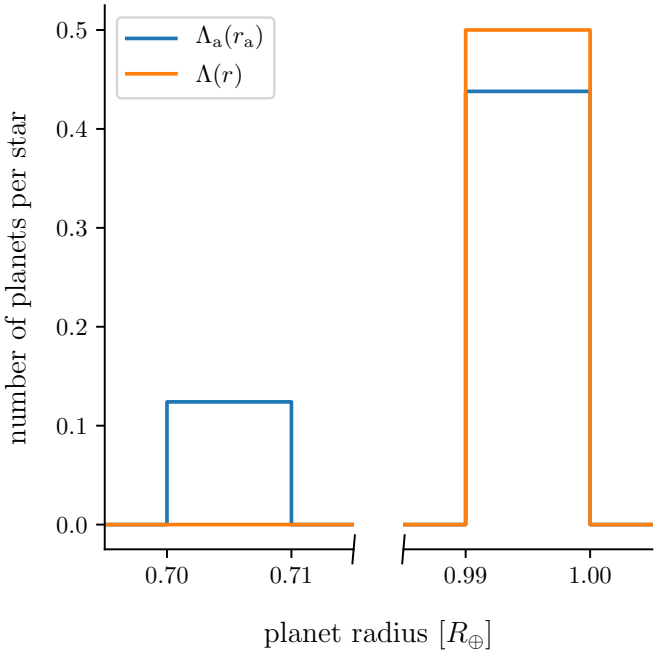

Figure 2. Apparent occurrence rate Λ_a, and true occurrence rate Λ in a universe with only one type of star, one type of planet (with radius $R_\oplus$), and a number density of binary stars 5% that of single stars. The occurrence rate of planets with radius $R_\oplus$ is underestimated, and the occurrence rate of planets with radius $R_\oplus/\sqrt{2}$ is overestimated.

which implies $n_b/n_s \approx 0.79$. Of course not all of these binaries are "twin" binaries as we have assumed in our simple calculation. Later, in Section 4, we will allow for a continuum of properties for the secondary stars. For now, we might guess that only a tenth of the binaries have pairs of stars close enough in brightness to produce errors as significant as those we have been considering. Adopting the value $n_b/n_s \approx 0.05$, we find

$$\frac{\Lambda_a(r_p)}{\Lambda(r_p)} = 0.88, \quad \frac{\Lambda_a(r_p/\sqrt{2})}{\Lambda(r_p)} = 0.25. \tag{9}$$

All together, the various effects produce biases of order 10% in the occurrence rates. We will see that this level of error is characteristic of many of our more complicated models as well.

3.2. *Formula Based on Apparent Rate Density*

There is a different way to conceptualize the preceding results that will be useful in the upcoming discussion. We now consider a spectrum of planet sizes, r, and introduce the occurrence rate density

$$\Gamma(r) \equiv \frac{d\Lambda}{dr}, \tag{10}$$

the number of planets per star per unit planet radius.

The naive observers are in fact measuring $\Gamma_a(r_a)$, the *apparent* occurrence rate density – the number of planets per star per unit apparent radius r_a. The "apparent

planet radius" is the radius inferred from the transit depth under the assumption that the host star is single. The "apparent stellar radius" $R_{\rm a}$ is analogous. The observed transit depth can be written $\delta = (r_{\rm a}/R_{\rm a})^2$, which from Equation 6 gives $r_{\rm a} = r_{\rm p}/\sqrt{2}$ for planets in twin binaries. $\Gamma_{\rm a}(r_{\rm a})$ is the same as the *true* rate density of transit signals of amplitude δ.

We will now calculate the apparent rate density by evaluating the star and planet counts for signals coming from single stars, primary stars in binaries, and secondary stars in binaries. Let μ be the ratio between the number of searchable binary systems and the number of searchable single systems at a particular transit depth. The number of single and binary sources in the searched sample are then

$$\frac{N_\star(r_{\rm a})}{1+\mu} \quad \text{and} \quad \frac{\mu N_\star(r_{\rm a})}{1+\mu}, \tag{11}$$

respectively. Earlier we showed that due to Malmquist bias, $\mu = 2^{3/2}(n_{\rm b}/n_{\rm s})$.

Suppose the occurrence rate densities for planets around single stars, primaries, and secondaries are $\Gamma_0(r), \Gamma_1(r)$, and $\Gamma_2(r)$, respectively. For a single planet size,

$$\Gamma_i(r) = \Lambda(r_{\rm p})\,\hat{\delta}(r - r_{\rm p}), \quad \text{for } i \in \{0,1,2\}, \tag{12}$$

where $\hat{\delta}$ is the Dirac delta function. The number of planets detected from singles, primaries, and secondaries are then

$$n^0_{\rm det}(r_{\rm a}){\rm d}r_{\rm a} = \frac{N_\star(r_{\rm a})}{1+\mu} p_{\rm tra}\Gamma_0(r_{\rm a}){\rm d}r_{\rm a}, \tag{13}$$

$$n^1_{\rm det}(r_{\rm a}){\rm d}r_{\rm a} = \frac{\mu N_\star(r_{\rm a})}{1+\mu} p_{\rm tra}\Gamma_1(\sqrt{2}r_{\rm a}){\rm d}(\sqrt{2}r_{\rm a}), \tag{14}$$

$$n^2_{\rm det}(r_{\rm a}){\rm d}r_{\rm a} = \frac{\mu N_\star(r_{\rm a})}{1+\mu} p_{\rm tra}\Gamma_2(\sqrt{2}r_{\rm a}){\rm d}(\sqrt{2}r_{\rm a}), \tag{15}$$

where $n^i_{\rm det}$ is the number of detections per unit apparent radius for each respective type of star.

Dividing the number of detections by the number of sources, the apparent rate density reported when ignoring binarity is

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{1}{1+\mu}\,\Gamma_0(r_{\rm a}) + \frac{\mu}{1+\mu}\,\sqrt{2}\,\Gamma_1(\sqrt{2}r_{\rm a}) + \frac{\mu}{1+\mu}\,\sqrt{2}\,\Gamma_2(\sqrt{2}r_{\rm a}). \tag{16}$$

If we insert Equation 12 and integrate over apparent radius, this reproduces Equations 4 and 7.

4. GENERAL FORMULA FOR APPARENT OCCURRENCE RATE

To generalize the procedure of Section 3.2 for binaries with varying light ratios, we consider an SNR-limited survey in which stars can have arbitrary properties. We assume that there are some functions $L(M)$ and $R(M)$ that specify a star's

luminosity and radius in terms of its mass. We also presume that the observers know the masses of every single and primary star, and that they assign each binary system the properties of its primary. Finally, we take the volume-limited distribution of binary mass ratios, $f(q)$, to be independent of the primary star's mass.

We want an equation for $\Gamma_{\rm a}(r_{\rm a}, M_{\rm a})$, the apparent rate density at each apparent planet radius and apparent stellar mass[1]. Similar to Equation 10, we define $\Gamma(r, M)$ as the number of planets per star, per unit planet radius, per unit stellar mass. Due to the varying binary mass ratios, we need the following modifications to Equation 16:

1. The Malmquist bias is different. The ratio between the number of binary and single systems for which a transit of depth δ is detectable, μ, has changed. At any $(r_{\rm a}, M_{\rm a})$ pair, the number of binary systems in the searchable volume with mass ratio $(q, q + {\rm d}q)$ is now

$$\frac{N_\star(r_{\rm a})}{1+\mu}\,\frac{n_{\rm b}}{n_{\rm s}}\left[\frac{L_{\rm tot}(M_{\rm a}, q)}{L(M_{\rm a})}\right]^{3/2} f(q)\,{\rm d}q, \tag{17}$$

   where μ is given by

$$\mu = \int_0^1 \frac{n_{\rm b}}{n_{\rm s}}\left[\frac{L_{\rm tot}(M_{\rm a}, q)}{L(M_{\rm a})}\right]^{3/2} f(q)\,{\rm d}q. \tag{18}$$

   Binaries with larger mass ratios are overrepresented by a factor of $[L_{\rm tot}(M_{\rm a}, q)/L(M_{\rm a})]^{3/2}$ because the sample of searchable binaries is magnitude-limited. To obtain the number of binary systems searched at each $(r_{\rm a}, M_{\rm a})$ pair, Equation 17 needs to be integrated over q. It can then replace $\mu N_\star(r_{\rm a})/(1+\mu)$ in Equations 14 and 15.

2. The apparent radius of a planet in a binary system now depends on whether the host star is the primary or the secondary. When the host is the primary, we write $r = \mathcal{D}_1 r_{\rm a}$, where

$$\mathcal{D}_1 = \left(\frac{L_{\rm tot}(M_{\rm a}, q)}{L(M_{\rm a})}\right)^{1/2}. \tag{19}$$

   When the host is the secondary, $r = \mathcal{D}_2 r_{\rm a}$, and the ratio between the assumed and true stellar radii must be included:

$$\mathcal{D}_2 = \frac{R(qM_{\rm a})}{R(M_{\rm a})}\left(\frac{L_{\rm tot}(M_{\rm a}, q)}{L(qM_{\rm a})}\right)^{1/2}. \tag{20}$$

3. Since the observers assign binaries the properties of the primary, the transit probability for planets orbiting secondaries is over-estimated. At a fixed orbital period, this changes $p_{\rm tra}$ in Equation 15 by a factor of

$$\frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3}. \tag{21}$$

[1] By assumption, the apparent stellar mass for single star systems is the true mass. For binary systems, the apparent mass is the mass of the primary.

4. The contribution to the apparent rate density from secondaries now depends on the mass of the secondary, qM_a. The associated measure on the right-hand side of Equation 15 becomes $\mathrm{d}(qM_\mathrm{a})$, and gives rise to an extra multiplicative factor of q.

Taking these modifications into account, a general formula for the apparent rate density follows:

$$\Gamma_\mathrm{a}(r_\mathrm{a}, M_\mathrm{a}) = \frac{1}{1+\mu}\left\{\Gamma_0(r_\mathrm{a}, M_\mathrm{a}) + \frac{n_\mathrm{b}}{n_\mathrm{s}}\left[\int_0^1 \mathrm{d}q\, \mathcal{D}_1^3 f(q)\cdot \mathcal{D}_1\Gamma_1(\mathcal{D}_1 r_\mathrm{a}, M_\mathrm{a}) + \int_0^1 \mathrm{d}q\, \mathcal{D}_1^3 f(q)\cdot q\mathcal{D}_2\Gamma_2(\mathcal{D}_2 r_\mathrm{a}, qM_\mathrm{a})\cdot \frac{R(qM_\mathrm{a})}{R(M_\mathrm{a})} q^{-1/3}\right]\right\}. \quad (22)$$

We give an alternative derivation in the Appendix. With the definition of μ in Equation 18, Equation 22 can also be expressed as

$$\Gamma_\mathrm{a}(r_\mathrm{a}, M_\mathrm{a}) = \frac{1}{1+\mu}\left[\Gamma_0(r_\mathrm{a}, M_\mathrm{a}) + \mu\left\langle \mathcal{D}_1\Gamma_1\left(\mathcal{D}_1 r_\mathrm{a}, M_\mathrm{a}\right) + q\mathcal{D}_2\Gamma_2\left(\mathcal{D}_2 r_\mathrm{a}, qM_\mathrm{a}\right)\cdot \frac{R(qM_\mathrm{a})}{R(M_\mathrm{a})} q^{-1/3}\right\rangle\right], \quad (23)$$

where the angle brackets denote averaging over $\mathcal{D}_1^3 f(q)$. Summarized, the apparent rate density is a weighted sum of the rate densities from singles, primaries, and secondaries. The weights are determined by the fraction μ and mass ratio distribution $\mathcal{D}_1^3 f(q)$ of binaries *in the searchable volume*. The fraction μ gives the relative contributions from singles and binaries, and $\mathcal{D}_1^3 f(q)$ applies a Malmquist bias for binaries that are brighter. A final correction for secondaries accounts for the overestimated transit probability.

5. REALISTIC STAR AND PLANET DISTRIBUTIONS

We will now apply our general equation for the apparent rate density (Equation 22) to study the impact of unrecognized binaries in regimes of observational interest. In the following, we write the rate density for each type of star, $\Gamma_i(r)$, as the product of a shape function, normalized to unity, and a constant:

$$\Gamma_i(r) = Z_i f_i(r), \quad \text{for } i \in \{0, 1, 2\}, \quad (24)$$

where the respective indices correspond to single stars, primaries of binaries, and secondaries of binaries. The normalizations Z_i are each system type's occurrence rate Λ_i, integrated over all planetary radii. In other words, they are number of planets per single, primary, or secondary star. Our main interest is in varying the radius shape function and the relative values of the integrated occurrence rates. We will also briefly mention shape functions in stellar mass. For analytic simplicity, we will assume throughout this section that stars are a one-parameter family, given by $L \propto M^\alpha \propto R^\alpha$. Applying these power laws to Equations 19 and 20, we find

$$\mathcal{D}_1 = (1+q^\alpha)^{1/2} \quad \text{and} \quad \mathcal{D}_2 = q(1+q^{-\alpha})^{1/2} \quad (25)$$

for the radius dilution prefactors.

5.1. *Power law planet radius distribution*

5.1.1. *Twin binaries*

As a first step, we return to the case where all binaries are twins. However, we now let the planet radius distribution be a power law,

$$\Gamma_i(r) = Z_i f(r) = Z_i r^\gamma / \mathcal{N}_r, \tag{26}$$

for $\mathcal{N}_r$ the shape function's normalization. Applying Equation 22 gives the resulting apparent rate density,

$$\Gamma_{\rm a}(r_{\rm a}) = \frac{r_{\rm a}^\gamma}{\mathcal{N}_r}\left[\frac{Z_0}{1+\mu} + 2^{\frac{\gamma+1}{2}}\frac{\mu}{1+\mu}(Z_1+Z_2)\right]. \tag{27}$$

Under these assumptions, twin binaries simply shift the apparent radius distribution: $\Gamma_{\rm a}(r_{\rm a}) \propto r_{\rm a}^\gamma$. As in Section 3.1, $\mu = 2^{3/2} n_{\rm b}/n_{\rm s}$. If planet occurrence is independent of system multiplicity, the "correction factor" relative to the rate density for singles is

$$\frac{\Gamma_{\rm a}(r_{\rm a})|_{r_{\rm a}=r}}{\Gamma_0(r)} = \frac{1+2^{\frac{\gamma+3}{2}}\mu}{1+\mu}. \tag{28}$$

For a binary fraction of 0.1, $\mu \approx 0.15$. Taking $\gamma = -2.92$ from Howard et al. (2012), we find that the apparent rate density is a mere 0.4% larger than the rate density of single stars. Fortuitously, the exponent of the radius distribution is close to -3, and the errors from neglecting binaries almost entirely cancel out.

The above estimate cannot apply at small radii, since the shape function diverges. Howard et al. (2012)'s results indicate that a single power law fits the planet radius distribution for radii greater than $2r_\oplus$. Similarly, there should be an upper radius limit, $r_{\rm u}$, to the radius distribution, perhaps around $24r_\oplus$ based on the most inflated hot Jupiters. Imposing an upper cut-off of $r_{\rm u}$ would lead to smaller values of the apparent rate density down to $r_{\rm u}/\sqrt{2}$, compared to a power law without the cut-off. We are not particularly interested in this effect because hot Jupiter occurrence rates are poorly described by a power law (see Section 5.3). All told, if our assumptions are applicable to real transit surveys, Equation 28 suggests that for apparent radii from $2r_\oplus$ to $17r_\oplus$, unrecognized binaries are a minor source of systematic error.

5.1.2. *Binaries with power law mass ratio distribution*

To check whether non-twin binaries change the preceding result, we now consider a distribution of binary mass ratios. Specifically, we take $f(q) = q^\beta/\mathcal{N}_q$, for $\mathcal{N}_q$ the normalization. This changes μ from the twin binary case; Equation 18 simplifies to

$$\mu = \frac{n_{\rm b}}{n_{\rm s}}\frac{1}{1+\beta}\int_0^1 (1+q^\alpha)^{3/2} q^\beta \, {\rm d}q. \tag{29}$$

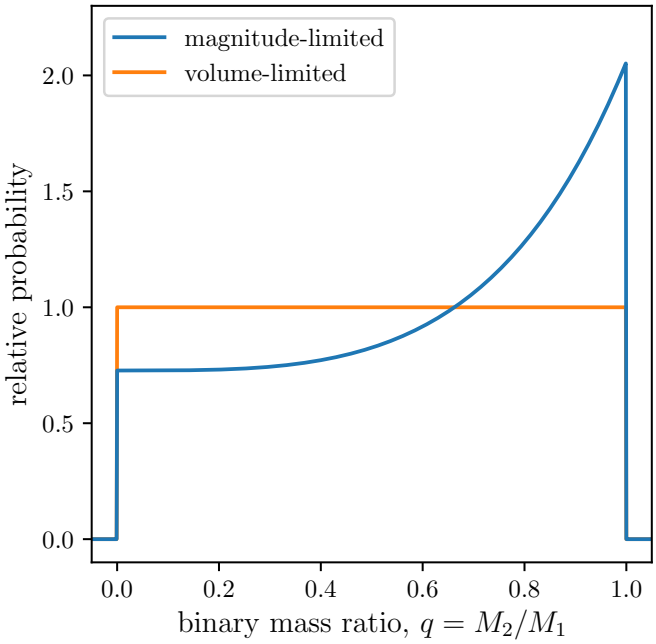

Figure 3. The mass ratio distribution for a magnitude-limited sample of binary stars, in which the underlying volume-limited distribution is uniform, qualitatively similar to Figure 16 of Raghavan et al. (2010). At a given observed transit depth, the searchable binaries in a transit survey are magnitude-limited. For this figure, we assume $L \propto M^{3.5}$.

For the $\beta = 0$ case, this equation means there will be more twin-like binaries than faint binaries in the sample of stars that are searched at any transit depth (Figure 3).

The rate density now varies not only with planet radius, but also host star mass. We can absorb this dependence into a power law,

$$\Gamma_i(r, M) = Z_i \frac{r^\gamma}{\mathcal{N}_r} \frac{M^\zeta}{\mathcal{N}_M}, \tag{30}$$

where Z_i is dimensionless, and the normalization constants carry the units.

Our main interest is in the apparent rate density's radius dependence. Marginalizing Equation 22 over apparent stellar mass, we find that when all stars host the same number of planets,

$$\frac{\Gamma_\mathrm{a}(r_\mathrm{a})|_{r_\mathrm{a}=r}}{\Gamma_0(r)} = \frac{1}{1+\mu}\left[1 + \frac{1}{\mathcal{N}_q}\frac{n_\mathrm{b}}{n_\mathrm{s}}\left(\int_0^1 \mathrm{d}q\, q^\beta (1+q^\alpha)^{\frac{\gamma+4}{2}} + \int_0^1 \mathrm{d}q\, q^{\beta+\gamma+\frac{5}{3}} (1+q^\alpha)^{\frac{3}{2}} (1+q^{-\alpha})^{\frac{\gamma+1}{2}}\right)\right], \tag{31}$$

where ζ does not appear because of the marginalization over M_a. For $\alpha = 3.5$, $\beta = 0$, $\gamma = -2.92$, the summed integrals in Equation 31 give $(\ldots) \approx 1.503$. For $n_\mathrm{b}/(n_\mathrm{b} + n_\mathrm{s}) = 0.44$, this yields $\Gamma_\mathrm{a}/\Gamma_0 = 1.048$. The apparent rate density is 4.8% larger than the rate density around singles. Considering only twin binaries gave us good intuition: for apparent radii from $2r_\oplus$ to $17r_\oplus$, binarity influences apparent planet occurrence rates around Sun-like stars at the level of a few percent.

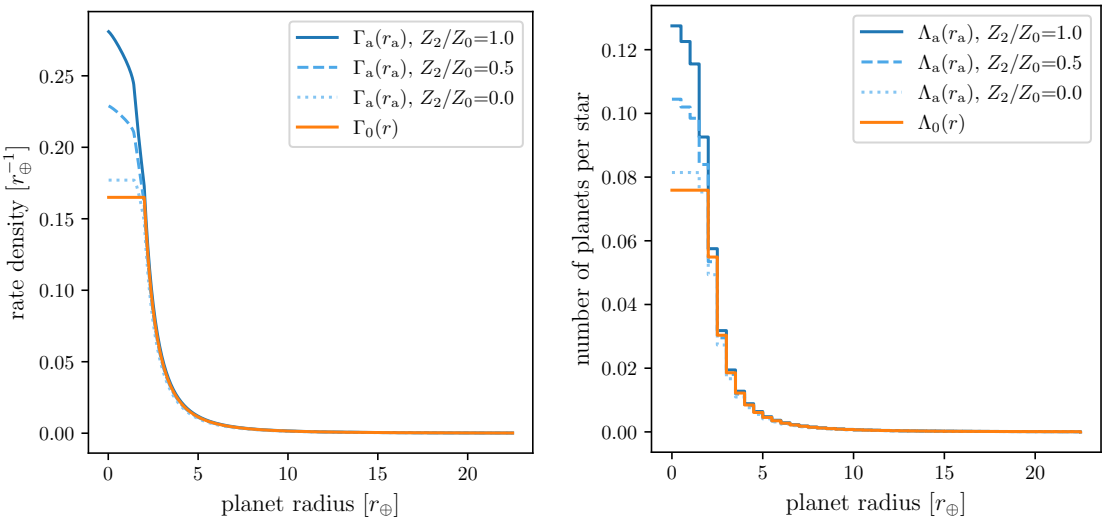

Figure 4. *Left:* apparent rate density (Γ_a) and rate density for single stars (Γ_0). *Right:* apparent occurrence rate (Λ_a) and single star occurrence rate (Λ_0), over $0.5r_\oplus$ bins. The true planet radius distribution is specified by Equation 32. This model assumes that the observer knows the true properties of all the singles and primaries, and that the volume-limited mass ratios of secondaries are drawn from a uniform distribution. Further, we take $Z_0 = Z_1 = 0.5$ throughout; Z_0, Z_1, Z_2 are the number of planets per single, primary, and secondary star. The rate and rate density are related by $\Lambda|_a^b = \int_a^b \Gamma \, dr$.

5.2. *Varying stars; broken power law planet radius distribution*

Though the details of the planet radius distribution $\Gamma_i(r)$ at $r < 2r_\oplus$ are currently an active topic of research, by making plausible assumptions we can investigate how neglecting binarity might bias measurements in this regime. For instance, consider a broken power-law:

$$f(r) \propto \begin{cases} r^\gamma & \text{for } r \geq 2r_\oplus \\ \text{constant} & \text{for } r \leq 2r_\oplus. \end{cases} \tag{32}$$

At apparent radii $r_a > 2r_\oplus$, the apparent rate density in this model is the same as that in Section 5.1.2. At smaller apparaent radii the equations are tedious, but still tractable. For simplicity, we insert Equation 32 into Equation 22, and integrate using a computer program. We refer the interested reader to our online implementation[2]. The output, shown in Figure 4, is validated using analytic predictions in the $r_a > 2r_\oplus$ and the $r_a < 2r_\oplus/\sqrt{2}$ regimes. As before, $L \propto M^\alpha \propto R^\alpha$, and $\alpha = 3.5$, $\beta = 0$, $\gamma = -2.92$.

The rate of Earth analogs—The immediately arresting result is a "bump" in the apparent rate density at apparent radii below $2r_\oplus$: the true rate for singles is less than the apparent rate. If secondaries host as many planets as single stars, the

[2] https://github.com/lgbouma/binary_biases, commit 6cb920c

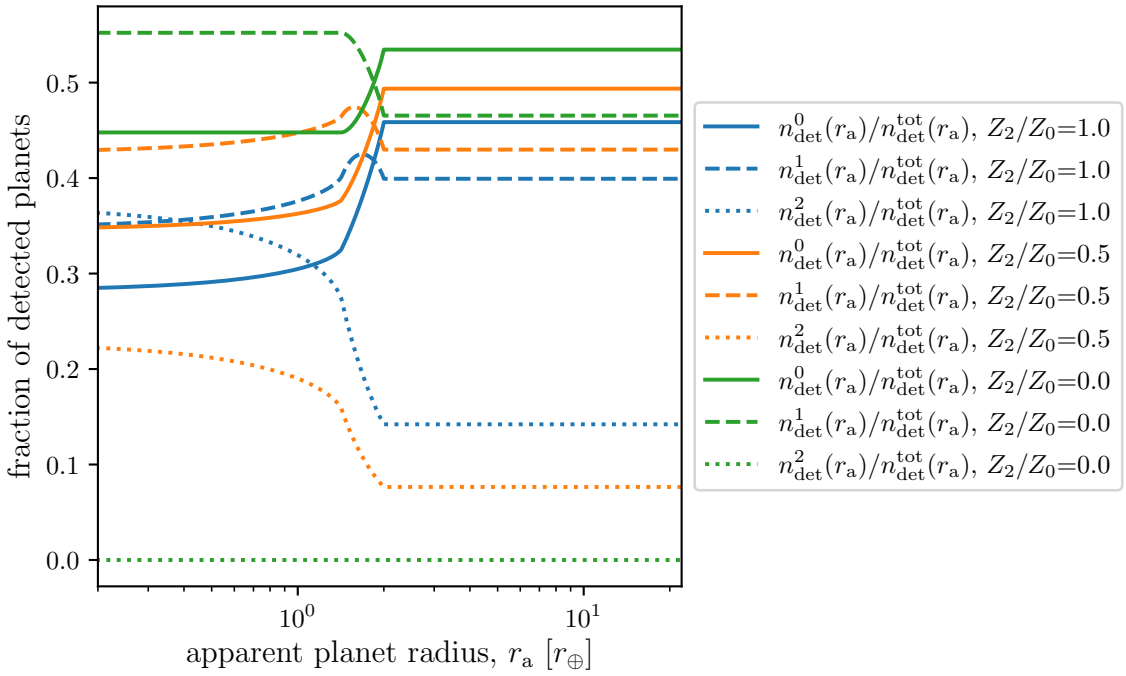

Figure 5. Fraction of detected planets at a given apparent planet radius that orbit single stars (solid lines), primaries of binaries (dashed lines), and secondaries of binaries (dotted lines). We assume that single stars and primaries have the same number of planets per star, and that the true planet radius distribution is a broken power law (same as Figure 4).

occurrence of these small planets is overestimated by about 50%. The magnitude of the systematic error decreases if there are fewer planets around secondaries; if secondaries host no planets at all, the apparent rate is overestimated by only 10%.

Hot Jupiter occurrence rates—By integrating the rate density from Figure 4, we can compare the apparent hot Jupiter occurrence rate with the true rate for singles. The difference between the rates is largest when secondaries host no planets. In that case, we find $\Lambda_{\rm HJ,0}/\Lambda_{\rm HJ,a} = 1.13$, where

$$\Lambda_{\rm HJ,a} = \int_{8r_\oplus}^{\infty} \Gamma_{\rm a}(r_{\rm a})\,{\rm d}r_{\rm a}, \quad \text{and} \quad \Lambda_{\rm HJ,0} = \int_{8r_\oplus}^{\infty} \Gamma_0(r)\,{\rm d}r. \tag{33}$$

If secondaries host half as many planets as single stars, then $\Lambda_{\rm HJ,0}/\Lambda_{\rm HJ,a} = 1.06$. We pay further attention to the issue of hot Jupiter occurrence rates in Section 5.3.

Fraction of detected planets from a given source—A common situation during transit survey follow-up is for high-resolution imaging to reveal multiple stars in a system that shows planetary transits. It is often impossible to rule which star hosts the detected planet. One would intuitively expect that the primary star would be the more likely host because at fixed planet size, planetary transits around secondaries have smaller amplitudes (Equations 19 and 20 imply $\mathcal{D}_2 > \mathcal{D}_1$, if $\alpha = 3.5$). However, this reasoning is qualitative, and its assumptions might not always hold.

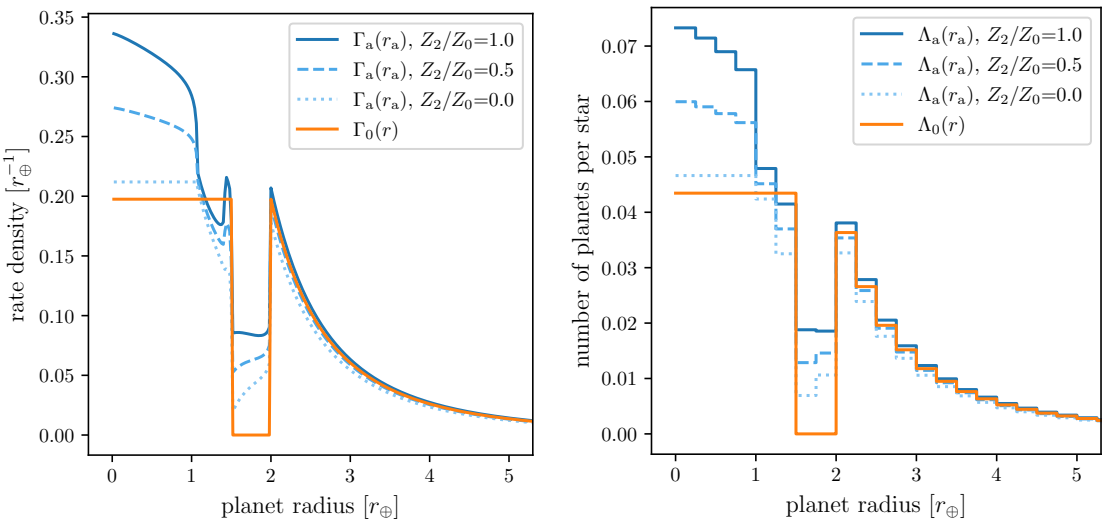

Figure 6. *Left:* rate density and *right:* rate as a function of planet radius in a model with a radius gap (Equation 34). Other than the intrinsic radius distribution, this model has the same assumptions as Figure 4.

Our formalism lets us calculate the relative probability that a detected planet orbits a single, primary, or secondary star in a more general manner. To perform the calculation, we use expressions given in the appendix for the number of detections coming from each type of system (Equations A18, A19, and A20). We then divide out by the total number of detections at each apparent radius. Figure 5 shows the result: if high-resolution imaging reveals a source with transits to be a binary system, the detected planet does usually orbit the primary.

If the apparent radius of the system is above $2r_\oplus$, and planets exist about primaries and secondaries at equal rates, then the planet is 3 times more likely to orbit the primary than the secondary. If secondaries host half as many planets as primaries, then a detected planet is 5 times more likely to orbit the primary.

The situation for apparent radii below $2r_\oplus$ is more nuanced. For the two cases in which secondaries host fewer planets than primaries, detected planets in binaries are still always more likely to orbit the primary. However, if planets exist at the same rate in primaries and secondaries, then below apparent radii of $0.4r_\oplus$, more of the detected planets in binaries come from secondaries – they have all been diluted down from larger true planetary radii. Rephrased, going from apparent radii of $2r_\oplus$ to $1.4r_\oplus$, the fraction of detected planets with binary companions increases by anywhere from 6% to 12%, depending on the relative occurrence of planets about primaries and secondaries. We compare this prediction with recent observations in Section 6.

5.3. *Further models: radius gaps, Gaussian HJ distributions*

The radius distribution specified by Equation 32 misses some important features revealed by state of the art planet occurrence studies.

Precise features of the radius valley—Fulton et al. (2017) recently reported a "gap" in the radius-period plane (Petigura et al. 2017a; Johnson et al. 2017). The existence of the gap has been independently corroborated from a sample of KOIs with asteroseismically-determined stellar parameters (Van Eylen et al. 2017). Precise measurement of the gap's features, in particular its width, depth, and shape, will require accurate occurrence rates. To illustrate binarity's role in this problem, we make identical assumptions as in Section 5.2, but instead assume an intrinsic radius distribution

$$f(r) \propto \begin{cases} r^{\gamma} & \text{for } r \geq 2r_{\oplus}, \\ 0 & \text{for } 1.5r_{\oplus} < r < 2r_{\oplus}, \\ \text{constant} & \text{for } r \leq 1.5r_{\oplus}. \end{cases} \tag{34}$$

Figure 6 shows the resulting true and apparent rates. If left uncorrected, binarity makes the gap appear more shallow, and flattens the step-function edges. Of course, effects unrelated to binarity would also "blur" the gap in the planet radius dimension. In particular, the valley's period-dependence is almost certainly not flat (Van Eylen et al. 2017; Owen & Wu 2017). This means that any study measuring the gap's width or depth while accounting for binarity must either perform tests at fixed orbital period, or else marginalize over period and account for the associated blurring in the planet radius dimension.

The HJ distribution is not a power law—In the recent study by Petigura et al. (2017b), hot Jupiters appear as an island in period-radius space, rather than as a continuous component of a power law distribution. This means that the apparent HJ rates computed in Section 5.2 are probably inaccurate. Instead, let us consider a Gaussian radius shape function,

$$f(r) \propto \exp\left(-\frac{(r-\bar{r})^2}{2\sigma^2}\right), \tag{35}$$

with $\bar{r} = 14r_{\oplus}$ and $\sigma = 2r_{\oplus}$. As always, $\Gamma_i(r) = Z_i f(r)$. We then compute $\Gamma_{\rm a}(r_{\rm a})$, and plot it in Figure 7. Integrating over $r_{\rm a} > 8r_{\oplus}$ to find hot Jupiter rates, we find the opposite effect as in the power law model. If secondaries host any planets at all, the apparent HJ rate is *greater* than the true rate for singles. For instance, if the hot Jupiter rate about secondaries and singles is equal, then $\Lambda_{\rm HJ,0}/\Lambda_{\rm HJ,a} = 0.81$. Qualitatively, the apparent HJ rate is greater than the true rate for singles because binary systems have extra uncounted stars that yield hot Jupiter detections.

6. DISCUSSION

How bad is ignoring binarity?—This study has shown that under a reasonable set of simplifying assumptions, ignoring binarity introduces systematic errors to star and planet counts in transit surveys, which then biases derived occurrence rates. Thus far, occurrence rate calculations[3] using transit survey data have mostly ignored stel-

[3] A list of occurrence rate papers is maintained at https://exoplanetarchive.ipac.caltech.edu/docs/occurrence_rate_papers.html

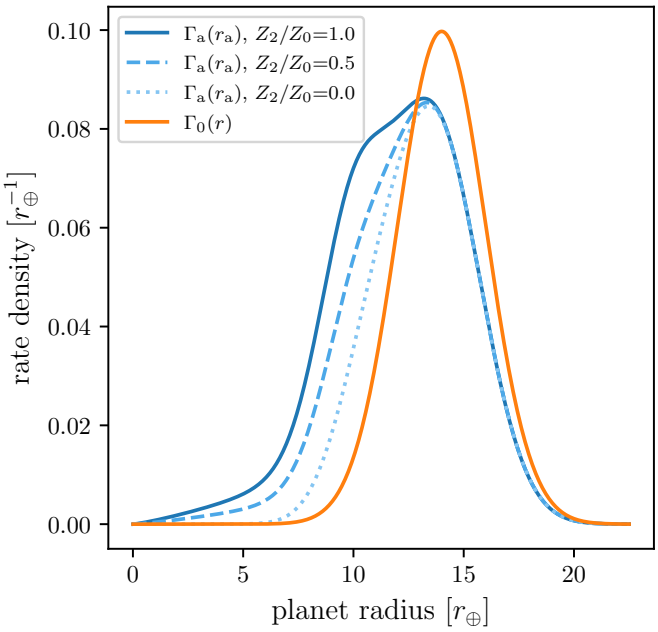

Figure 7. Rate density for a population of planets with true radii r drawn from a Gaussian with mean $14r_\oplus$ and standard deviation $2r_\oplus$. This is similar to the hot Jupiter distribution presented by Petigura et al. (2017b).

lar multiplicity (*e.g.*, Howard et al. 2012; Fressin et al. 2013; Foreman-Mackey et al. 2014; Dressing & Charbonneau 2015; Burke et al. 2015). For *Kepler* occurrence rates specifically, it seems that no one has yet carefully assessed binarity's importance, or lack thereof. This study does not resolve the problem; it only suggests the approximate scale of the systematic errors. Section 5.1 suggests that for apparent radii above $2r_\oplus$, binarity can be ignored down to a precision of a few percent. For apparent radii below $2r_\oplus$, the picture is less forgiving: Section 5.2 suggests that the apparent rates around single stars could be overestimated by as much as 50%.

The rate of Earth analogs—Youdin (2011), Petigura et al. (2013), Dong & Zhu (2013), Foreman-Mackey et al. (2014), and Burke et al. (2015) have found that between 0.03 and 1 Earth-like planets exist per Sun-like star, depending on assumptions that are made when calculating the rate (see Burke et al. 2015, Figure 17). Assuming a broken power-law, we showed that unrecognized binaries could cause overestimates in the rate of Earth analogs of up to 50%. This systematic effect is smaller than the other factors that currently dominate the dispersion in $\eta_\oplus$ measurements. If future analyses determine absolute values of $\eta_\oplus$ to better than a factor of two, binarity will likely merit closer attention.

One caveat is that none of our models included the rate density's period-dependence. However, binaries with separations $\lesssim 10\,\mathrm{AU}$ could provoke dynamical instabilities, leading to fewer Earth-like planets per star (*e.g.*, Holman & Wiegert 1999; Wang et al. 2014; Kraus et al. 2016). This would affect transit survey measurements of $\eta_\oplus$ beyond our rough estimate.

Hot Jupiter rate discrepancy—While unresolved binaries may bias $\eta_\oplus$ measurements, our approach suggests that they do not influence the hot Jupiter rate discrepancy. The "discrepancy" is that RV surveys find roughly twice as many hot Jupiters per Sun-like star as do transit surveys (roughly 1% vs. 0.5%, respectively). We summarize hot Jupiter rates reported by different surveys in Table 1.

Though the disagreement is only weakly significant ($< 3\sigma$), one reason to expect a difference is that the RV sample is biased towards metal-rich stars, which host more giant planets (Santos et al. 2004; Fischer & Valenti 2005; Gould et al. 2006). Investigating the discrepancy from the metallicty angle, Guo et al. (2017) measured the *Kepler* field's mean metallicity to be $[\mathrm{M/H}]_{\mathrm{Kepler}} = -0.045 \pm 0.009$, which is lower than the California Planet Search's mean of $[\mathrm{M/H}]_{\mathrm{CPS}} = -0.005 \pm 0.006$. The former value agrees with that found by Dong et al. (2014). Refitting for the metallicity exponent in $\Lambda_{\mathrm{HJ}} \propto 10^{\beta[\mathrm{M/H}]}$, Guo et al. found $\beta = 2.1 \pm 0.7$, and noted that the metallicity difference could account for 0.2% of the needed 0.5% difference between the measured CKS and *Kepler* rates[4]. Guo et al. concluded that "other factors, such as binary contamination and imperfect stellar properties" must also be at play.

Radial velocity surveys usually reject visual and spectroscopic binaries (Wright et al. 2012), so their hot Jupiter rates are closer to the rate for single stars than the rates reported by transit surveys. However, we found in Section 5.3 that when assuming a Gaussian radius distribution, apparent hot Jupiter rates are *greater* than the true rate around singles: the effect goes the wrong way. There could be other systematic factors behind the difference – but they are unlikely to be related to binarity.

Does a detected planet orbit the primary or secondary?—A separate reason to address stellar multiplicity in transit surveys is that it can radically alter the interpretation of planet candidates on a system-by-system level. To identify and correct for unseen stellar companions, high-resolution imaging campaigns have surveyed almost all *Kepler* Objects of Interest (Howell et al. 2011; Adams et al. 2012, 2013; Horch et al. 2012, 2014; Lillo-Box et al. 2012, 2014; Dressing et al. 2014; Law et al. 2014; Cartier et al. 2015; Everett et al. 2015; Gilliland et al. 2015; Wang et al. 2015b,c; Baranec et al. 2016; Ziegler et al. 2017). The results of these programs have been summarized by Furlan et al. (2017), and they represent an important advance in understanding the KOI sample's multiplicity statistics.

An immediate application is to reassess the radii of detected planets. Doing so, Hirsch et al. (2017) recently found that for their sample of KOIs in binaries, the planet radii are underestimated by factors $r/r_{\mathrm{a}} = 1.17$ if all planets orbit primaries, and $r/r_{\mathrm{a}} = 1.65$ if detected planets are equally likely to orbit primaries and singles. This approach makes minimal assumptions about the likelihood that a planet orbits

[4] Petigura et al. (2017b) recently found $\beta = 3.4^{0.9}_{-0.8}$. If true, this implies that metallicity could account for about half of the hot Jupiter rate discrepancy.

the primary vs. the secondary (though Hirsch et al. 2017 do also consider weighting radius corrections by planet occurrence).

Our approach provides a suggestion of how to assess the probability that a detected planet in a binary system orbits either component. The problem requires assuming both the underlying radius distribution, and the relative number of planets per primary and secondary star. We found that for a broken power law radius distribution, with equal occurrence rates about primaries and secondaries, detected planets in binary systems are usually more likely to orbit the primary (Figure 5). For apparent radii greater than $2r_\oplus$, the odds of primary:secondary are 8:3. The odds increase for the planet to orbit the secondary increase at smaller apparent radii, and reach 1:1 at $r_a = 0.4r_\oplus$. For *Kepler*, this suggests that for almost all confirmed KOIs in multiple star systems, the planet most likely orbits the primary.

Fraction of detected planets with binary companions—A related effect is that some detected planets are more likely to be in binary systems than others. Notably, Figure 5 predicts that the fraction of detected planets with binary companions should increase by 6 to 12% going from $2r_\oplus$ to $1.4r_\oplus$.

In Ziegler et al. (2017)'s recent summary of the Robo-AO KOI survey, they reported the fraction of planet-hosting stars with Robo-AO detected companion stars, binned over apparent planet radii. The fraction of KOIs with nearby stars and their 1σ uncertainties are:

- Earths ($r_a < 1.6r_\oplus$): $16.3 \pm 1.0\%$, from 1480 systems.
- Neptunes ($1.6r_\oplus < r_a < 3.9r_\oplus$): $13.0 \pm 0.8\%$, from 2058 systems.
- Saturns ($3.9r_\oplus < r_a < 9r_\oplus$): $13.6 \pm 2.0\%$, from 338 systems.
- Jupiters ($r_a > 9r_\oplus$): $19.0 \pm 2.8\%$, from 247 systems.

The uncertainties were calculated following Burgasser et al. (2003). The absolute values of the companion fractions are lower than in the solar neighborhood at least in part because of Robo-AO's sensitivity (Ziegler et al. 2017's Figure 2; Raghavan et al. 2010). The uptick in the companion rate for Jupiters is likely tied to a large astrophysical false positive rate (Santerne et al. 2012). Going from Neptunes to Earths, the data suggest a weak increase in the detected planet companion fraction, perhaps of a few percent. It will be interesting to see whether this effect is borne out by further observations.

Connection between high-resolution imaging and occurrence rates—Occurrence rate calculations are beginning to incorporate the results of high-resolution imaging surveys. For example, recent studies have reduced contamination in the numerator of the occurrence rate by using Furlan et al. (2017)'s catalog to test the effects of removing KOI hosts with known companions (Fulton et al. 2017; Petigura et al. 2017b). This is a good first step, but the denominator remains uncorrected.

Even if the multiplicity of every star that *Kepler* surveyed were known, the problem would not be solved. In this scenario, it would be possible to tabulate planet occurrence for single and binary star systems separately. The number of planets per single star would be known. However the number of planets per primary and secondary would still be convolved. An observational approach to separating the two populations might be to analyze the centroids, or the transit durations (and thus host star densities) for a representative sample of planets in binary systems.

A theory-driven solution might entail probabilistic population inference, with a parametrized model for planet populations around primary and secondary stars. Such an approach might be expected to find that the primary and secondary populations are different, since single M dwarfs host more small planets than FGK dwarfs (Dressing & Charbonneau 2015). Given the difficulty of writing a likelihood function in a survey with systematic biases, one possible approach might be to forward-model the survey in the framework of approximate Bayesian computation (*e.g.*, Morehead 2016).

7. CONCLUSION

This study presented a framework for estimating the impact of unresolved binaries on transit survey occurrence rates. From the order-of-magnitude argument presented in Section 3, we showed that for Sun-like stars in the local neighborhood, typical twin binary fractions yield apparent occurrence rates that are well within a factor of two of the true rates for single stars.

We then derived a general formula for the apparent rate density inferred by observers who ignore binarity (Equation 22). As input, this equation requires a volume-limited mass ratio distribution, and the true rate densities for planets around singles, primaries, and secondaries. The assumptions that enable this approach are as follows.

1. The transit survey is SNR-limited (Equation 1).

2. The observers know the true properties of the single and primary stars, and assign all unresolved binaries the properties of the primary.

3. There are two functions, $L(M)$ and $R(M)$, that specify a star's luminosity and radius in terms of its mass.

The interpretation of Equation 22 is that the apparent rate density is a weighted sum of the rate densities for single, primary, and secondary stars. The weights depend on the relative numbers of binary and single stars in the searchable volumes, which are affected by both dilution of transit signals and a Malmquist bias. Both effects should be considered in Monte Carlo simulations of transit surveys (*e.g.*, Bakos et al. 2013; Sullivan et al. 2015; Günther et al. 2017).

Applying Equation 22, we then showed that:

- If single, primary, and secondary stars all host the same number of planets per star, and the planet radius distribution is a power law with the exponent

reported by Howard et al. (2012), then binarity influences apparent planet occurrence rates at the few percent level. This applies for apparent radii from $2r_\oplus$ to $17r_\oplus$ (Section 5.1).

- Assuming a broken-power law planet radius distribution, with the Howard et al. (2012) exponent above $2r_\oplus$ and a constant occurrence below $2r_\oplus$, the rate density for radii below $2r_\oplus$ is overestimated. The magnitude of the error, $\Delta\Gamma_0 = |\Gamma_0 - \Gamma_a|/\Gamma_0$, can easily reach 50% (Figure 4). Although this is smaller than current systematic uncertainties on the occurrence rates of Earth-sized planets, this implies that binarity could eventually become an important component of the $\eta_\oplus$ error budget.

- Binarity blurs sharp features in the radius distribution (Figure 6), to a degree that could affect precise measurements of the depth, width, and slope of the Fulton et al. (2017) radius gap.

- Binarity does not lead to smaller apparent HJ occurrence rates (Figure 7). This assumes a Gaussian planet radius distribution, similar to that reported by Petigura et al. (2017b).

- Detected planets with apparent radii greater than $0.5r_\oplus$ that are revealed by high-resolution imaging surveys to exist in binaries are more likely to orbit the primary (Figure 5).

- Near the break in occurrence as a function of planet radius at $2r_\oplus$, the fraction of detected planets with binary companions should increase by roughly 5 to 10% (Figure 5).

These results are only strictly applicable for idealized transit surveys meeting the criteria mentioned above. For real transit surveys, although our approach is only suggestive, we hope that it provides a useful estimate for the systematic errors that can be incurred by ignoring binarity in calculations of planetary occurrence rates.

It was a pleasure discussing this study with T. Barclay, W. Bhatti, J. Christiansen, F. Dai, and T. Morton. This work made use of NASA's Astrophysics Data System Bibliographic Services.

Software: `numpy` (Walt et al. 2011), `scipy` (Jones et al. 2001), `matplotlib` (Hunter 2007), `pandas` (McKinney 2010), `IPython` (Pérez & Granger 2007)

APPENDIX

A. ALTERNATIVE DERIVATION OF APPARENT RATE DENSITY

This appendix derives Equation 22 through a more straight-forward, but also more tedious, procedure than that of Section 4. First, note that analogous to Equation 2, the apparent rate density $\Gamma_{\rm a}$ can be written

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{n_{\rm det}(r_{\rm a}, M_{\rm a})}{N_\star(r_{\rm a}, M_{\rm a})} \frac{1}{p_{\rm tra}(M_{\rm a})}, \tag{A1}$$

where $n_{\rm det}$ is the number of detected planets, per unit $(r_{\rm a}, M_{\rm a})$, with apparent radius $r_{\rm a}$ orbiting stars of apparent mass $M_{\rm a}$. The planets with $(r_{\rm a}, M_{\rm a})$ are associated with systems of many different planetary and stellar properties, so $n_{\rm det}(r_{\rm a}, M_{\rm a})$ is given by the convolution of the true rate density, $\Gamma(r, M)$, and $\mathcal{N}_\star(r_{\rm a}, M_{\rm a}; r, M)$, the number of searchable stars per unit $(r_{\rm a}, M_{\rm a})$ that give $(r_{\rm a}, M_{\rm a})$ when the true system actually has properties (r, M). Mathematically,

$$n_{\rm det}(r_{\rm a}, M_{\rm a}) = \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}) \tag{A2}$$

$$= \sum_i \int {\rm d}r{\rm d}M\, \mathcal{N}^i_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_i(r, M)\, p_{\rm tra}(r, M), \tag{A3}$$

where i specifies the type of true host star (0: single, 1: primary, 2: secondary). The problem reduces to the evaluation of

$$\mathcal{N}^i_\star(\mathcal{P}_{\rm a}, \mathcal{S}_{\rm a}; \mathcal{P}, \mathcal{S}) \tag{A4}$$

for planets around single stars, primaries in binaries, and secondaries in binaries.

Single stars—For $i = 0$,

$$\mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M) = \hat{\delta}(r_{\rm a} - r)\hat{\delta}(M_{\rm a} - M) N^0_\star(r, M), \tag{A5}$$

where $\hat{\delta}$ is the Dirac delta function, so

$$n^0_{\rm det}(r_{\rm a}, M_{\rm a}) = N^0_\star(r_{\rm a}, M_{\rm a})\, \Gamma_0(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(r_{\rm a}, M_{\rm a}). \tag{A6}$$

Primaries in binaries—The number of primaries with apparent parameters $(r_{\rm a}, M_{\rm a})$ given the true parameters (r, M) is

$$\mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M) = \int {\rm d}q\, f(q) \mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M), \tag{A7}$$

where $f(q)$ is the volume-limited binary mass ratio distribution. Since we assume that binaries are assigned the mass of the primary,

$$\mathcal{N}^1_{{\rm s},q}(r_{\rm a}, M_{\rm a}, q; r, M) \propto \hat{\delta}(M_{\rm a} - M). \tag{A8}$$

In this case, $\mathcal{N}^1_{\mathrm{s},q}$ is non-zero only at $r_\mathrm{a} = R_\mathrm{a}\sqrt{\delta}$, and the observed depth is

$$\delta = \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{L(M_\mathrm{a})}{L_\mathrm{tot}(M_\mathrm{a}, q)} \equiv \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{1}{\mathcal{D}_1^2} \tag{A9}$$

where

$$\mathcal{D}_1 = \sqrt{\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(M_\mathrm{a})}}. \tag{A10}$$

The normalization of $\mathcal{N}^1_{\mathrm{s},q}$ is given by the number of binaries that are searchable for a signal δ and that have mass ratio q:

$$N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \frac{n_\mathrm{b}}{n_\mathrm{s}} \left[\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(M_\mathrm{a})}\right]^{3/2}. \tag{A11}$$

where the argument for the number of searchable single stars, $N^0_\star$, could also be expressed as $(\delta, L(M_\mathrm{a}))$. Applying Equation A11,

$$\mathcal{N}^1_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) = N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \frac{n_\mathrm{b}}{n_\mathrm{s}} \mathcal{D}_1^3 \hat{\delta} \left(r_\mathrm{a} - \frac{r}{\mathcal{D}_1}\right) \hat{\delta}(M_\mathrm{a} - M). \tag{A12}$$

Secondaries in binaries—In this case, $M = qM_1 = qM_\mathrm{a}$, so

$$\mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) \propto \hat{\delta} \left(M_\mathrm{a} - \frac{M}{q}\right). \tag{A13}$$

Again $\mathcal{N}^2_\star$ is non-zero only at $r_\mathrm{a} = R_\mathrm{a}\sqrt{\delta}$, but this time

$$\delta = \left[\frac{r}{R(qM_\mathrm{a})}\right]^2 \frac{L(qM_\mathrm{a})}{L_\mathrm{tot}(M_\mathrm{a}, q)} \equiv \left[\frac{r}{R(M_\mathrm{a})}\right]^2 \frac{1}{\mathcal{D}_2^2}, \tag{A14}$$

where

$$\mathcal{D}_2 = \frac{R(qM_\mathrm{a})}{R(M_\mathrm{a})} \sqrt{\frac{L_\mathrm{tot}(M_\mathrm{a}, q)}{L(qM_\mathrm{a})}}. \tag{A15}$$

The normalization remains the same as the previous case – we are counting the searchable stars at a given observed depth δ, and the total luminosity of the binary is the same. Thus,

$$\mathcal{N}^2_\star(r_\mathrm{a}, M_\mathrm{a}; r, M) = \int \mathrm{d}q\, f(q) \mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M), \tag{A16}$$

where

$$\mathcal{N}^2_{\mathrm{s},q}(r_\mathrm{a}, M_\mathrm{a}, q; r, M) = N^0_\star(r_\mathrm{a}, M_\mathrm{a}) \frac{n_\mathrm{b}}{n_\mathrm{s}} \mathcal{D}_1^3 \hat{\delta} \left(r_\mathrm{a} - \frac{r}{\mathcal{D}_2}\right) \hat{\delta} \left(M_\mathrm{a} - \frac{M}{q}\right). \tag{A17}$$

One might worry in Equation A17 that we opt to write $\mathcal{N}^2_\mathrm{s} \propto \hat{\delta}(M_\mathrm{a} - M/q)$, rather than $\propto \hat{\delta}(M_\mathrm{a} q - M)$ or some other delta function with the same functional dependence, but a different normalization once integrated. We do this because the delta function in Equation A17 is defined with respect to the measure $\mathrm{d}M_\mathrm{a}$, not $\mathrm{d}M$. This is because $\mathcal{N}^2_\mathrm{s}$ is defined as a number per r_a, per M_a.

Number of detected planets—Marginalizing per Equation A3, we find

$$\begin{aligned} n^0_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^0_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_0(r, M)\, p_{\rm tra}(M) \\ &= N^0_\star(r_{\rm a}, M_{\rm a})\, \Gamma_0(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(M_{\rm a}), \end{aligned} \tag{A18}$$

and

$$\begin{aligned} n^1_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^1_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_1(r, M)\, p_{\rm tra}(M) \\ &= N^0_\star(r_{\rm a}, M_{\rm a})\, p_{\rm tra}(M_{\rm a})\, \frac{n_{\rm b}}{n_{\rm s}} \int {\rm d}q\, \mathcal{D}_1^3 f(q)\, \mathcal{D}_1 \Gamma_1(\mathcal{D}_1 r_{\rm a}, M_{\rm a}). \end{aligned} \tag{A19}$$

Finally,

$$\begin{aligned} n^2_{\rm det}(r_{\rm a}, M_{\rm a}) &= \int {\rm d}r{\rm d}M\, \mathcal{N}^2_\star(r_{\rm a}, M_{\rm a}; r, M)\, \Gamma_2(r, M)\, p_{\rm tra}(M) \\ &= N^0_\star(r_{\rm a}, M_{\rm a})\, \frac{n_{\rm b}}{n_{\rm s}} \int {\rm d}q\, \mathcal{D}_1^3 f(q)\, q\mathcal{D}_2 \Gamma_2(\mathcal{D}_2 r_{\rm a}, qM_{\rm a})\, \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3}. \end{aligned} \tag{A20}$$

General formula for apparent occurrence rate—Combining the above results with Equation 11, the apparent rate density,

$$\Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{N_\star(r_{\rm a}, M_{\rm a}) p_{\rm tra}(r_{\rm a}, M_{\rm a})} \sum_i n^i_{\rm det}(r_{\rm a}, M_{\rm a}), \tag{A21}$$

evaluates to

$$\begin{aligned} \Gamma_{\rm a}(r_{\rm a}, M_{\rm a}) = \frac{1}{1+\mu} \Bigg\{ \Gamma_0(r_{\rm a}, M_{\rm a}) + \frac{n_{\rm b}}{n_{\rm s}} \Bigg[&\int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q)\, \mathcal{D}_1 \Gamma_1(\mathcal{D}_1 r_{\rm a}, M_{\rm a}) \\ &+ \int_0^1 {\rm d}q\, \mathcal{D}_1^3 f(q)\, q\mathcal{D}_2 \Gamma_2(\mathcal{D}_2 r_{\rm a}, qM_{\rm a})\, \frac{R(qM_{\rm a})}{R(M_{\rm a})} q^{-1/3} \Bigg] \Bigg\}. \end{aligned} \tag{A22}$$

We validate this equation in limits where it is possible to write down the answer (e.g., Equation 16), and also against a Monte Carlo realization of the twin binary models (Sections 3.1 and 5.1).

Table 1. Occurrence rates of hot Jupiters (HJs) about FGK dwarfs, as measured by radial velocity and transit surveys.

Reference	HJs per thousand stars	HJ Definition
Marcy et al. (2005)	12±2	$a < 0.1$ AU; $P \lesssim 10$ day
Cumming et al. (2008)	15±6	–
Mayor et al. (2011)	8.9±3.6	–
Wright et al. (2012)	12.0±3.8	–
Gould et al. (2006)	$3.1^{+4.3}_{-1.8}$	$P < 5$ day
Bayliss & Sackett (2011)	10^{+27}_{-8}	$P < 10$ day
Howard et al. (2012)	4±1	$P < 10$ day; $r_p = 8 - 32 r_\oplus$; solar subset[a]
–	5±1	solar subset extended to $Kp < 16$
–	7.6±1.3	solar subset extended to $r_p > 5.6 r_\oplus$.
Moutou et al. (2013)	10±3	*CoRoT* average; $P \lesssim 10$ day, $r_p > 4 r_\oplus$
Petigura et al. (2017b)	$5.7^{+1.4}_{-1.2}$	$r_p = 8 - 24 r_\oplus$; $P = 1 - 10$ day; CKS stars[b]
Santerne et al. (2018, in prep)	9.5±2.6	*CoRoT* galactic center
–	11.2±3.1	*CoRoT* anti-center

NOTE— The first four studies use data from radial velocity surveys; the rest are based on transit surveys. Many of these surveys selected different stellar samples. "–" denotes "same as above".

[a] Howard et al. (2012)'s "solar subset" was defined as *Kepler*-observed stars with $4100\,\mathrm{K} < T_{\rm eff} < 6100\,\mathrm{K}$, $Kp < 15$, $4.0 < \log g < 4.9$. They required signal to noise > 10 for planet detection.

[b] Petigura et al. (2017b)'s planet sample includes all KOIs with $Kp < 14.2$, with a statistically insignificant number of fainter stars with HZ planets and multiple transiting planets. Their stellar sample begins with Mathur et al. (2017)'s catalog of 199991 *Kepler*-observed stars. Successive cuts are: $Kp < 14.2$ mag, $T_{\rm eff} = 4700 - 6500$ K, and $\log g = 3.9 - 5.0$ dex, leaving 33020 stars.

REFERENCES

Adams, E. R., Ciardi, D. R., Dupree, A. K., et al. 2012, The Astronomical Journal, 144, 42

Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2013, The Astronomical Journal, 146, 9

Bakos, G. ., Csubry, Z., Penev, K., et al. 2013, PASP, 125, 154

Baranec, C., Ziegler, C., Law, N. M., et al. 2016, The Astronomical Journal, 152, 18

Bayliss, D. D. R., & Sackett, P. D. 2011, The Astrophysical Journal, 743, 103

Burgasser, A. J., Kirkpatrick, J. D., Reid, I. N., et al. 2003, The Astrophysical Journal, 586, 512

Burke, C. J., Christiansen, J. L., Mullally, F., et al. 2015, The Astrophysical Journal, 809, 8

Cartier, K. M. S., Gilliland, R. L., Wright, J. T., & Ciardi, D. R. 2015, The Astrophysical Journal, 804, 97

Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, Publications of the Astronomical Society of the Pacific, 120, 531

Dong, S., & Zhu, Z. 2013, The Astrophysical Journal, 778, 53

Dong, S., Zheng, Z., Zhu, Z., et al. 2014, The Astrophysical Journal Letters, 789, L3

Dressing, C. D., Adams, E. R., Dupree, A. K., Kulesa, C., & McCarthy, D. 2014, The Astronomical Journal, 148, 78

Dressing, C. D., & Charbonneau, D. 2015, ApJ, 807, 45

Everett, M. E., Barclay, T., Ciardi, D. R., et al. 2015, The Astronomical Journal, 149, 55

Fischer, D. A., & Valenti, J. 2005, The Astrophysical Journal, 622, 1102

Foreman-Mackey, D., Hogg, D. W., & Morton, T. D. 2014, The Astrophysical Journal, 795, 64

Fressin, F., Torres, G., Charbonneau, D., et al. 2013, The Astrophysical Journal, 766, 81

Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, The Astronomical Journal, 154, 109

Furlan, E., Ciardi, D. R., Everett, M. E., et al. 2017, The Astronomical Journal, 153, 71

Gilliland, R. L., Cartier, K. M. S., Adams, E. R., et al. 2015, The Astronomical Journal, 149, 24

Gould, A., Dorsher, S., Gaudi, B. S., & Udalski, A. 2006, Acta Astronomica, 56, 1

Günther, M. N., Queloz, D., Demory, B.-O., & Bouchy, F. 2017, Monthly Notices of the Royal Astronomical Society, 465, 3379

Guo, X., Johnson, J. A., Mann, A. W., et al. 2017, The Astrophysical Journal, 838, 25

Hirsch, L. A., Ciardi, D. R., Howard, A. W., et al. 2017, The Astronomical Journal, 153, 117

Holman, M. J., & Wiegert, P. A. 1999, The Astronomical Journal, 117, 621

Horch, E. P., Howell, S. B., Everett, M. E., & Ciardi, D. R. 2012, The Astronomical Journal, 144, 165

—. 2014, The Astrophysical Journal, 795, 60

Howard, A. W., Marcy, G. W., Bryson, S. T., et al. 2012, The Astrophysical Journal Supplement Series, 201, 15

Howell, S. B., Everett, M. E., Sherry, W., Horch, E., & Ciardi, D. R. 2011, The Astronomical Journal, 142, 19

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90

Johnson, J. A., Petigura, E. A., Fulton, B. J., et al. 2017, arXiv:1703.10402 [astro-ph]

Jones, E., Oliphant, T., Peterson, P., et al. 2001, Open source scientific tools for Python

Kraus, A. L., Ireland, M. J., Huber, D., Mann, A. W., & Dupuy, T. J. 2016, The Astronomical Journal, 152, 8

Law, N. M., Morton, T., Baranec, C., et al. 2014, The Astrophysical Journal, 791, 35

Lillo-Box, J., Barrado, D., & Bouy, H. 2012, Astronomy and Astrophysics, 546, A10

—. 2014, Astronomy and Astrophysics, 566, A103

Marcy, G., Butler, R. P., Fischer, D., et al. 2005, Progress of Theoretical Physics Supplement, 158, 24

Mathur, S., Huber, D., Batalha, N. M., et al. 2017, The Astrophysical Journal Supplement Series, 229, 30

Mayor, M., Marmier, M., Lovis, C., et al. 2011, ArXiv e-prints, 1109, arXiv:1109.2497

McKinney, W. 2010, in Proceedings of the 9th Python in Science Conference, ed. S. van der Walt & J. Millman, 51

Morehead, R. C. 2016, Ph.D. Thesis

Moutou, C., Deleuil, M., Guillot, T., et al. 2013, Icarus, 226, 1625

Owen, J. E., & Wu, Y. 2017, arXiv:1705.10810 [astro-ph]

Pepper, J., & Gaudi, B. S. 2005, The Astrophysical Journal, 631, 581

Pepper, J., Gould, A., & Depoy, D. L. 2003, Acta Astronomica, 53, 213

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Petigura, E. A., Howard, A. W., & Marcy, G. W. 2013, Proceedings of the National Academy of Science, 110, 19273

Petigura, E. A., Howard, A. W., Marcy, G. W., et al. 2017a, arXiv:1703.10400 [astro-ph]

Petigura, E. A., Marcy, G. W., Winn, J. N., et al. 2017b, arXiv:1712.04042 [astro-ph]

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, The Astrophysical Journal Supplement Series, 190, 1

Santerne, A., Daz, R. F., Moutou, C., et al. 2012, Astronomy and Astrophysics, 545, A76

Santos, N. C., Israelian, G., & Mayor, M. 2004, Astronomy and Astrophysics, 415, 1153

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., et al. 2015, The Astrophysical Journal, 809, 77

Van Eylen, V., Agentoft, C., Lundkvist, M. S., et al. 2017, arXiv:1710.05398 [astro-ph]

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. 2011, Computing in Science & Engineering, 13, 22

Wang, J., Fischer, D. A., Horch, E. P., & Huang, X. 2015a, The Astrophysical Journal, 799, 229

Wang, J., Fischer, D. A., Horch, E. P., & Xie, J.-W. 2015b, The Astrophysical Journal, 806, 248

Wang, J., Fischer, D. A., Xie, J.-W., & Ciardi, D. R. 2015c, The Astrophysical Journal, 813, 130

Wang, J., Xie, J.-W., Barclay, T., & Fischer, D. A. 2014, The Astrophysical Journal, 783, 4

Winn, J. N. 2010, Exoplanet Transits and Occultations, ed. S. Seager (University of Arizona Press), 55

Wright, J. T., Marcy, G. W., Howard, A. W., et al. 2012, The Astrophysical Journal, 753, 160

Youdin, A. N. 2011, ApJ, 742, 38

Ziegler, C., Law, N. M., Baranec, C., et al. 2017, arXiv:1712.04454 [astro-ph]