Draft version October 28, 2017
Typeset using LaTeX **modern** style in AASTeX61

THE EFFECTS OF BINARITY ON PLANET OCCURRENCE RATES MEASURED BY TRANSIT SURVEYS

L. G. Bouma, J. N. Winn

ABSTRACT

This work aims to clarify the biases that stellar binarity introduces to occurrence rates inferred from transit surveys. In general, stellar multiplicity leads to diluted planetary radii, overestimated detection efficiencies, and an undercounted number of selected stars (though possibly an overcounted number of searched stars). These effects skew occurrence rate measurements in different directions, and we develop simple models that allow us to understand the crucial effects. For models in which all stellar systems are either single or twin binaries, and all planets are identical, we find that ignoring binarity leads to a two-to-threefold underestimate of the occurrence at the true radius. We believe this represents an upper bound on binarity's effects. Using more realistic models for the stellar population and planetary radii, we find that ignoring binarity leads to a 10 - 30% underestimate of the number of planets per star, depending on planet radii. For the occurrence of Earth-sized planets, in our most realistic model the underestimate is by $\approx 15\%$ – at present far smaller than other systematic uncertainties. For hot Jupiters, we find that the inferred occurrence rate is $\approx 1.3\times$ smaller than the true rate around single stars. We suggest that this latter effect contributes to the discrepant hot Jupiter rates measured by *Kepler* and the California group's respective surveys.

1. INTRODUCTION

An astronomer who does not believe in stellar multiplicity wants to measure the mean number of planets of a certain type per star of a certain type. They perform a signal-to-noise limited transit survey and detect $N_{\rm det}$ transit signals that appear to be from planets of the desired type. They calculate the apparent number of searchable stars, $N_\star$. "Searchable stars" are the stars for which planets are observed with 100% detection efficiency. Correcting for the geometric transit probability $f_{\rm g}$, they compute an apparent occurrence rate $\Lambda_{\rm a}$:

$$\Lambda_{\rm a} = \frac{N_{\rm det}}{N_\star} \times \frac{1}{f_g}. \tag{1}$$

There are many potential pitfalls. Some genuine transit signals can be missed by the detection pipeline. Some apparent transit signals are spurious, from noise fluctuations, failures of 'detrending', or instrumental effects. Stars and planets can be misclassified due to statistical and systematic errors in the measurements of their properties. Poor angular resolution causes false positives due to blends with background eclipsing binaries. *Et cetera.*

Here we focus on problems that arise from the fact that many stars exist in multiple star systems. For simplicity, we only consider binaries, and we assume that they are all spatially unresolved.

An immediate complication is that, due to dynamical stability or some aspect of planet formation, the occurrence rate of planets may differ between binary and single-star systems. If "occurrence rate" is defined purely as the mean number of planets within set radius and period bounds per star in a given mass interval, it must implicitly marginalize over stellar multiplicity. This means marginalizing over "occurrence rates in single star systems", "occurrence rates about primaries", and "occurrence rates about secondaries" (see *e.g.,* Wang et al., 2015).

Outside of astrophysical differences, there are observational biases. A given apparently-searchable star may truly be a single star of the desired type. If it is binary, it may contain two, one, or no searchable stars of the desired type, which affects $N_\star$. There may be systematic errors in stellar parameter estimates of such systems, which in this work we neglect. One might also imagine an apparently-searchable star which in no stellar component is of the desired type. We will not consider errors of this type. We will assume that all the apparently-searchable stars are either single stars of the desired type, or binaries in which either the primary or else both components are of the desired type.

When counting the detected signals from planets of the desired type, $N_{\rm det}$, ignoring binarity leads to errors in planet radius classification, and the assumed survey completeness. Detected planets in binary systems will have underestimated radii because of the diluting flux from the companions, and possibly because they are assumed to orbit the wrong star (*e.g.*, Furlan et al. 2017). The number of selected stars should

change dependent on the planet radius and period for which the rate measurement is desired, and on the survey's achieved photometric precision. This mean our naive astronomer assumes the fraction of their selected stars which are searchable is 1 (this is their "assumed completeness"). Binarity confuses this, because the ability to search for planets in binary systems depends on which stellar component hosts the planet, and on how much dilution affects the measurement.

How Has Binarity Been Considered in Occurrence Rate Measurements?—Many authors have reported planet occurrence rates from transit survey data[1]. With regard to binarity, the typical approach has been to ignore the issue[2]. For instance, in a laudable study by Burke et al. (2015), they

> "treat the pipeline completeness as having no uncertainty due to the incomplete understanding of the stellar parameters, eccentricity, and stellar binarity."

Though they do "explore sensitivity in the derived planet occurrence rates to alternative assumptions for the stellar parameters and non-zero eccentricities", they do not for binarity.

Of course, on a system-by-system level stellar multiplicity affects the interpretation of planet candidates. High resolution imaging campaigns have consequently been undertaken to measure the multiplicity of 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. 2015a, 2015b; Baranec et al. 2016). The results of these programs have been collected by Furlan et al. (2017), and they represent an important advance in understanding the KOI sample's multiplicity statistics. In particular, they can be immediately applied to rectify binarity's influence on the mass-radius diagram (Furlan & Howell 2017).

For *Kepler*-derived occurrence rates, the high resolution imaging campaigns have not yet come full circle to observe a comparison sample of non-KOI host stars. The most recent rate studies have thus used Furlan et al. (2017)'s catalog to test the effects of removing KOI hosts with known companions, which is an important step towards reducing contamination in the "numerator" of the occurrence rate (Fulton et al. 2017). However, without an understanding of the multiplicity statistics of the non-KOI hosting stars that are assumed to be searchable, the true completeness, the true number of searchable stars, and thus the true occurrence rates will remain uncertain.

The Hot Jupiter Rate Discrepancy—There is at least one context in which measurement of absolute occurrence rates may already be showing the signatures of binarity.

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

[2] Deacon et al., 2015 are an exception to the rule.

Hot Jupiter occurrence rates measured by transit surveys ($\approx 0.5\%$) are marginally lower than those found by radial velocity surveys ($\approx 1\%$; see Table 1). Though the discrepancy is of weak statistical significance ($< 3\sigma$), one plausible explanation for the difference is that the populations have distinct metallicities. As originally argued by Gould et al. (2006), the RV sample is biased towards metal-rich stars, which have been measured by RV surveys to preferentially host more giant planets (Santos et al 2004, Fischer and Valenti 2005). The *Kepler* sample specifically has been measured to be more metal poor than the local neighborhood, with a mean metallicity of $[\mathrm{M/H}]_{\mathrm{mean}} \approx -0.05$ (Dong et al., 2014; Guo et al., 2017). Studying the problem in detail, Guo et al. recently argued that the metallicity difference could account for a $\approx 0.1\%$ difference in the measured rates between the CKS and *Kepler* samples – not a $\approx 0.5\%$ difference. Guo et al. concluded that "other factors, such as binary contamination and imperfect stellar properties" must also be at play.

Aside from surveying stars of varying metallicities, radial velocity and transit surveys differ in how they treat binarity. Radial velocity surveys typically reject both visual and spectroscopic binaries (*e.g.*, Wright et al. 2012). Transit surveys typically observe binaries, but the question of whether they were searchable to begin with is usually left for later interpretation. In spectroscopic follow-up of candidate transiting planets, the prevalence of astrophysical false-positives may also lead to a bias against confirmation of transiting planets in binary systems.

Ignoring these complications, in this work we focus on whether binarity might intrinsically bias transit survey occurrence rates, simply through its effects on the number of searchable stars and the apparent radii of detected planets.

To progressively gain intuition, we study the following idealized transit surveys:

- Model #1: fixed stars, fixed planets, twin binaries;
- Model #2: fixed planets and primaries, varying secondaries;
- Model #3: fixed primaries, varying planets and secondaries.

In Sec. 2.1, we introduce our numerical approach to the problem, and in Sec. 2.2 we clarify our terminology. We present the analytic and numerical results in Secs. 3.1-3.3, where each subsection corresponds to each model above. We interpret these calculations throughout, and in Sec. 4 discuss their relevance to various questions in the interpretation of transit survey occurrence rates.

2. METHOD

2.1. *Numerical Approach*

Assuming a signal-to-noise limited survey, we would like to find the true occurrence rate density, and also that inferred by an observer ignoring binarity. Though analytic or semi-analytic solutions exist for each model, beyond Model #2 the equations become burdensome. To maintain simplicity, we develop a Monte Carlo approach, which works as follows.

First, the user specifies the model class (#1, 2, or 3) and various free parameters describing the stellar and planetary population. Most importantly, these parameters include the binary fraction, and the true planet occurrence rates around single stars, primaries, and secondaries (the Λ_i's, to be defined in Sec. 2.2).

We then generate the population of selected stars. Each selected star has a type (single, primary, secondary), a binary mass ratio (if it is not single), and the property of whether it is "searchable". The absolute number of stars is arbitrary. The relative number of binaries to singles is calculated according to analytic formulae. The binary mass ratios are sampled from the appropriate magnitude-limited distributions.

Whether a star is "searchable" depends entirely on its "completeness" fraction. By "completeness", we mean the ratio of the actual number of searchable stars to the assumed number of searchable stars (for a given planet size, period, etc.). Assuming homogeneously distributed stars, this is equivalent to the ratio of the searchable to selected volumes. For single stars, we assume these volumes are identical – exactly the case discussed by Pepper et al. (2003). For binaries, this volume ratio is a function of only the binary mass ratio.

To assign planets, each selected star receives a planet at the initially specified rate, according to its type. The radii of planets are assigned independently of any host system property, as sampled from $p_r(r) \sim r^\delta$ for Model #3 or a δ–function for Models #1 and #2. A planet is detected when a) it transits, and b) its host star is searchable.

The probability of transiting single stars in our model is assumed to be known, and so it is mostly corrected by the observer attempting to infer an occurrence rate. The only bias is for secondaries, which can be smaller than primaries in Models #2 and #3. This effect is included when computing the transit probability.

For detected planets, apparent radii are computed according to analytic formulae that account for both dilution and the misclassification of stellar radii. We assume that the observer think all transits are around single stars.

The rates are then computed in bins of true planet radius and apparent planet radius. In a given radius bin, the true rate is found by counting the number of planets that exist around selected stars of all types (singles, primaries, secondaries), and dividing by the total number of these stars. The apparent rates are found by counting the number of detected planets that were found in an apparent radius bin, dividing by the geometric transit probability for single stars, and dividing by the apparent total number of stars. The simplest realization of this scheme is described analytically in Sec. 3.1, but we first clarify our terminology.

2.2. *Analytic Preliminaries*

Define the occurrence rate density, Γ, as the expected number of planets per star per natural logarithmic bin of planetary and stellar phase space:

$$\Gamma(\vec{x}) = \frac{d^n \Lambda}{\prod_{i=1}^{n} \mathrm{d} \ln x_i}. \tag{2}$$

$\vec{x}$ is an n-dimensional list of the continuous physical properties that might affect the occurrence rate density. For example, $\vec{x} = (r, P, R)$ where r is the planet radius, P is its orbital period, and R is the host star radius. The occurrence rate Λ is found by integrating the rate density over a specified volume of phase space (e.g., Youdin 2011).

The previous definition implicitly marginalizes the rate density over stellar multiplicity. For simplicity, this work only considers single and binary star systems. Then for a selected population of stars and planets the rate density can be written

$$\Gamma(\vec{x}) = \sum_{i=0}^{2} w_i \Gamma_i(\vec{x}) = \sum_i w_i \Lambda_i p_i(\vec{x}) \tag{3}$$

where $i = 0$ corresponds to single star systems, $i = 1$ primaries of binaries, and $i = 2$ secondaries of binaries. Λ_i is the occurrence rate integrated over all possible phase space for the i^{th} system type, and $p_i(\vec{x})$ is the joint probability density function so that $\Gamma_i(\vec{x}) = \Lambda_i p_i(\vec{x})$. The weights are given by

$$w_i = N_i / N_{\text{tot}}, \tag{4}$$

for $N_{\text{tot}} = \sum_i N_i$ the total number of selected stars, and N_0, N_1, N_2 the number of selected single stars, primaries, and secondaries respectively[3]. The relationship between the rate Λ over a desired volume of phase space Ω_{desired} and Λ_i is

$$\Lambda = \sum_i \left[\left(w_i \Lambda_i \int_{\Omega_{\text{desired}}} p_i(\vec{x})\, \mathrm{d}\Omega \right) \cdot \left(w_i \Lambda_i \int_{\Omega_{\text{total}}} p_i(\vec{x})\, \mathrm{d}\Omega \right)^{-1} \right], \tag{5}$$

where the inverse term is unity if $p_i(\vec{x})$ is appropriately normalized.

A transit survey will have a rate density of detected planets $\hat{\Gamma}$, which for each system type will be the product of the rate density and the detection efficiency $Q_i(\vec{x})$:

$$\hat{\Gamma}(\vec{x}) = \sum_i Q_i(\vec{x}) \Gamma_i(\vec{x}) \equiv \sum_i \hat{\Gamma}_i(\vec{x}), \tag{6}$$

where again the index i is over each type of system (singles, primaries, and secondaries). The detection efficiency includes the geometric transit probability, as well as any incompleteness effects. Foreman-Mackey et al. (2014) discuss how this is calculated in practice.

3. ANALYTIC AND NUMERICAL RESULTS

3.1. *Model #1: fixed stars, fixed planets, twin binaries*

[3] Eq. 3 follows by writing the i^{th} system type's rate density as some normalization multiplied by a probability density: $\Gamma_i(\vec{x}) = \mathcal{Z}_i p_i(\vec{x})$. For Eq. 2 to hold, we must have $\mathcal{Z}_i = \Lambda_i$.

Consider a universe in which all planets are identical, and all stars are either single or twin stars with otherwise identical physical properties. The occurrence rate density at a planet radius r, stellar radius R, and semimajor axis a for this model is

$$\Gamma(r, R, a) = \sum_i w_i \Lambda_i \delta^3(r_p, R_\star, a_p), \tag{7}$$

for r_p and $R_\star$ some fixed planet and stellar radii, a_p a fixed semi-major axis, and δ the Dirac delta function, whose compact form will be used for brevity. The occurrence rate over any interval that includes r_p, $R_\star$, and a_p is

$$\Lambda = \sum_i w_i \Lambda_i = \frac{\sum_i N_i \Lambda_i}{N_{\rm tot}}. \tag{8}$$

The rate is zero over intervals that do not.

To express the rate density of detected planets, $\hat{\Gamma} = \sum Q_i \Gamma_i$, we need the detection efficiencies for each system type, which are products of the geometric and selection probabilities:

$$Q_i(\vec{x}) = Q_{g,i}(\vec{x}) Q_{c,i}(\vec{x}), \quad \text{where } \vec{x} = (r, R, a). \tag{9}$$

Similar to Pepper et al. (2003), but in a new context, we take Q_c as the ratio of the number of stars that were searchable to the number of stars that were selected. Assuming a homogeneous distribution of stars, this gives

$$Q_{c,i}(\vec{x}) = \left(\frac{d_{{\rm det},i}(\vec{x})}{d_{\rm sel}(\vec{x})} \right)^3, \tag{10}$$

for $d_{\rm sel}$ the maximum distance to which surveyed stars are selected, and $d_{{\rm det},i}$ the maximum distance to which planets can actually be detected about the $i^{\rm th}$ system type. Note that $d_{\rm sel} \geq d_{{\rm det},i}$. In a signal-to-noise limited transit survey in which the observer does not know which stars are binaries,

$$d_{\rm sel} \propto (r/R)^2 (L_{\rm sys} T_{\rm dur} A N_{\rm tra})^{1/2}, \tag{11}$$

for $L_{\rm sys} = L_1(1 + \ell)$ the system luminosity, $T_{\rm dur}$ the transit duration, A the detector area, and $N_{\rm tra}$ the number of observed transits. However,

$$d_{{\rm det},i} \propto \mathcal{D}_i (r/R)^2 (L_{\rm sys} T_{\rm dur} A N_{\rm tra})^{1/2}, \tag{12}$$

for the dilution $\mathcal{D}_i$ given by

$$\mathcal{D}_i = \begin{cases} 1 & \text{for } i = 0, \text{ single} \\ L_1/L_{\rm sys} = (1 + \ell)^{-1}, & \text{for } i = 1, \text{ primary} \\ \ell L_1/L_{\rm sys} = (1 + \ell^{-1})^{-1}, & \text{for } i = 2, \text{ secondary,} \end{cases} \tag{13}$$

where the light ratio ℓ of a given binary is defined as the ratio of the luminosity of the secondary to the primary.

The maximum detectable distance to single stars is assumed to be known, and so $d_{\rm sel,0} = d_{\rm det,0}$. For binary systems there is a necessary incompleteness, and combining Eqs. 9 through 13 yields

$$Q_0 = Q_{g,0} Q_{c,0} = Q_{g,0} \tag{14}$$
$$Q_1 = Q_{g,1} Q_{c,1} = Q_{g,0}(1+q^{\alpha})^{-3} \tag{15}$$
$$Q_2 = Q_{g,2} Q_{c,2} = Q_{g,0} q^{2/3}(1+q^{-\alpha})^{-3} q^{-5}, \tag{16}$$

for $Q_{g,0} = R/a$, the transit probability in single star systems. The factors of $q^{2/3}$ and q^{-5} in Eq. 16 come from the assumed stellar mass-luminosity-radius relation: $R \propto M \propto L^{1/\alpha}$. For $q = 1$ both terms evaluate to unity, but they will later become relevant.

Summarizing, we have written the rate density for each system type (Eq. 7) and the detection efficiency for each system type (Eq. 14-16), and so have fully specified the rate density of detected planets, in addition to the true rate density.

What does an observer ignoring binarity infer?—An observer who ignores binarity assumes a detection efficiency $\tilde{Q} = Q_0$, measures a detected planet rate density $\tilde{\Gamma}$, and infers an apparent rate density Γ_a. Analogous to Eq. 6,

$$\tilde{\Gamma} = \Gamma_a \tilde{Q}. \tag{17}$$

Accounting for dilution, one can show

$$\Gamma_a = w_a \Lambda_0 \delta^3(r_p, R_\star, a_p) + w_b(\Lambda_1 Q_{c,1} + \Lambda_2 Q_{c,2})\delta^3(r_p/\sqrt{2}, R_\star, a_p), \tag{18}$$

for $w_a = N_0/(N_0+N_1)$, and $w_b = N_1/(N_0+N_1)$. This observer miscounts the number of total searched stars, does not correct for incompleteness, and misclassifies the planetary radii because of dilution.

Correction to inferred rate density and inferred rate—Define a rate density correction factor, X_Γ, as the ratio of the apparent to true rate densities:

$$X_\Gamma \equiv \frac{\Gamma_a}{\Gamma}. \tag{19}$$

This factor can be a function of whatever parameters Γ_a and Γ depend on; in this study, the planet radius is most relevant. For the twin-binaries model,

$$X_\Gamma(r) = \frac{w_a \Lambda_0 \delta^3(r_p) + w_b(\Lambda_1 Q_{c,1} + \Lambda_2 Q_{c,2})\delta^3(r_p/\sqrt{2})}{(w_0\Lambda_0 + w_1\Lambda_1 + w_2\Lambda_2)\delta^3(r_p)} \tag{20}$$

where $\delta^3(r_p)$ is shorthand for $\delta^3(r - r_p, R - R_\star, a - a_p)$.

If we take the rates Λ_i to be equal, applying the definitions of the weights gives a rate density correction factor at $r = r_p$ of $X_\Gamma(r_p) = (1+\mu)^{-1}$, where

$$\mu \equiv \frac{N_1}{N_0} = \frac{n_b}{n_s}\left(\frac{d_{\rm sel,b}}{d_{\rm sel,s}}\right)^3 = \frac{\rm BF}{1-\rm BF}(1+\ell)^{3/2}, \tag{21}$$

for n_b and n_s the number density of binaries and singles in a volume limited sample. Using Raghavan et al. (2010)'s $0.7-1.3M_\odot$ multiplicity fraction as our binary fraction[4], we set BF = 0.44. The resulting correction to the rate density is $X_\Gamma(r_p) \approx 0.31$. The correction at $r_p/\sqrt{2}$ is infinite. The numerical realization of this model agrees with these analytic values, and its output is shown in Fig 1. If instead we assume that $\Lambda_0 = \Lambda_1$, but that $\Lambda_2 = 0$, we find $X_\Gamma(r_p) = (1+2\mu)/(1+\mu)^2$. Taking the same binary fraction, this evaluates to $X_\Gamma(r_p) \approx 0.53$. Since the correction to the rate is equal to that of rate density, at $r = r_p$, the occurrence rate is underestimated by a factor of roughly 2 to 3.

3.2. *Model #2: fixed planets and primaries, varying secondaries*

The main use of our binary-twin model is to help develop intuition. We now let the light ratio $\ell = L_2/L_1$ vary across the binary population. It does so because the underlying mass ratio $q = M_2/M_1$ varies. We keep the primary mass fixed as M_1, which is also the mass of all single stars.

We parametrize the distribution of binary mass ratios in a volume-limited sample as a power law: $p(q) \propto q^\beta$. For binaries with solar-type primaries[5], β is probably between 0 and 0.3. Since we assume stars are a one-parameter family, $R \propto M \propto L^{1/\alpha}$, a drawn value of q determines everything about a secondary.

The rate density in this model, $\Gamma(\vec{x})$, is the sum over system types of $w_i\Lambda_i p_i(\vec{x})$:

$$\Gamma(\vec{x}) = \delta^4(r_p, R_\star, a_p, P_p)(w_0\Lambda_0 + w_1\Lambda_1) + w_2\Lambda_2\delta^3(r_p, P_p, a_p)p_2(q), \tag{22}$$

where the semimajor axis of the planet must be such that its period is P_p, and $p_2(q)$ is expressed in terms of the mass ratio instead of the secondary star's radius for convenience (q and R_2 are interchangeable). The probability that a secondary hosts a planet, as a function of the mass ratio, is

$$p_2(q) = p(\text{has planet} \,|\, \text{secondary with } q) \times p(\text{secondary with } q) \tag{23}$$

$$p_2(q) \propto q^{\gamma+\beta}(1+q^\alpha)^{3/2}. \tag{24}$$

We take first term, $p(\text{has planet} \,|\, \text{secondary with } q)$, as a power law of q with exponent γ. For the second term, since the selected sample at a given (r, P, a) is magnitude-limited, $p(\text{secondary with } q)$ is the product of the volume limited probability, q^β, and

[4] The binary fraction is the fraction of systems in a volume-limited sample that are binary. It is equivalent to the multiplicity fraction if there are no triple, quadruple, or higher order multiples.

[5] Duchene and Kraus (2013), fitting all the multiple systems of Raghavan et al. (2010)'s Fig 16, find $\beta = 0.28 \pm 0.05$ for $0.7 < M_\star/M_\odot < 1.3$. Examining only the binary systems of Rhagavan et al 2010, Fig 16, the distribution seems roughly uniform, $\beta \approx 0$, except for a claimed excess of twin binaries with $q \approx 1$, and an obvious lack of $q < 0.1$ stellar companions.

a Malmquist-like bias, $(1+q^{\alpha})^{3/2}$. We note that various authors (probably ourselves included) have incorrectly drawn from volume-limited binary distributions in Monte Carlo simulations of transit surveys (Sullivan et al 2015, Bakos et al 2012, Guenther et al 2017, Bouma et al 2017). The correct distribution for a magnitude-limited selection is shown in Fig. 2.

The occurrence rate corresponding to Eq. 22's rate density for a desired volume of phase space $\Omega_{\rm desired}$ is given by Eq. 5. Specifying the desired mass ratios of interest as $q_{\rm min} < q < q_{\rm max}$, this simplifies to

$$\Lambda = \frac{N_0\Lambda_0 + N_1\Lambda_1 + N_2\Lambda_2 f_2}{N_{\rm tot}}, \tag{25}$$

for

$$f_2 \equiv \left(\int_{q_{\rm min}}^{q_{\rm max}} p_2(q)\,{\rm d}q\right) \cdot \left(\int_0^1 p_2(q)\,{\rm d}q\right)^{-1}. \tag{26}$$

The detected rate density, $\hat{\Gamma} = \sum_i Q_i \Gamma_i$, will be specified by the detection efficiencies for each type of system. These are identical to Eqs. 14-16. The detection efficiency for secondaries (Eq. 16) includes the transit probability from the smaller stellar radius, and combines dilution, the transit duration, and stellar radius for the completeness probability.

What does an observer ignoring binarity infer?—As a reminder, the apparent rate density is found by correcting the detected apparent rate density for the transit probability: $\Gamma_a = \tilde{\Gamma} Q_{g,0}^{-1}$. The observer's errors are as follows:

1. The true planetary radii r are interpreted as apparent radii r_a. The apparent radii depend on the system type:

$$r_a = \begin{cases} r_p(1+q^{\alpha})^{-1/2} & \text{for } i=1,\ \text{primary} \\ r_p(1+q^{-\alpha})^{-1/2} q^{-1}, & \text{for } i=2,\ \text{secondary.} \end{cases} \tag{27}$$

   The factor of q^{-1} for the secondary case accounts for the observer assuming that the host star is the primary.

2. The completeness fraction is miscalculated for any stars in binary systems.

3. The selected and searchable stars are miscounted.

To write the apparent rate density as a function of the apparent radius r_a, we marginalize out the planet period, semimajor axis, and stellar radius (or equivalently the mass ratio, for binaries):

$$\Gamma_a(r_a) = w_a\Lambda_0\delta(r_p) + w_b\Lambda_1 I_1(r_a) + w_b\Lambda_2 I_2(r_a), \tag{28}$$

where the detection efficiencies are given in Eqs. 14-16, and as in the first model, $w_a = N_0/(N_0+N_1)$, $w_b = N_1/(N_0+N_1)$. The ratio of primaries to singles, μ, is now

given by a variant of Eq. 21:

$$\mu \equiv \frac{N_1}{N_0} = \frac{\mathrm{BF}}{1-\mathrm{BF}}\left(2^{3/2} - \int_1^{\sqrt{2}} u^2(u^2-1)^{1/\alpha}\,\mathrm{d}u\right), \tag{29}$$

where the latter dimensionless integral is easily found numerically. The $I_1(r_a)$ and $I_2(r_a)$ terms marginalize over the joint distribution of apparent radius and mass ratio:

$$I_i(r_a) = \int_0^1 p(\text{has detected planet}, r_a, q|\text{star is type } i)\,\mathrm{d}q, \quad \text{for } i \in \{1,2\}, \tag{30}$$

$$= \int_0^1 p(\text{has detected planet}|r_a, q, \text{star is type } i) \times p(r_a|q, \text{star is type } i)\, p(q|\text{star is type } i)\,\mathrm{d}q. \tag{31}$$

The first term is the detection efficiency; the second is a δ-function of the apparent radius; the last is the mass ratio distribution given by Eq. 24. The analytic solution for $i = 1$ is

$$I_1(r_a) = \frac{1}{\mathcal{N}_1}\left(\frac{r_p}{r_a}\right)^{-3}\left(\left(\frac{r_p}{r_a}\right)^2 - 1\right)^{\frac{\gamma+\beta}{\alpha}}, \quad \text{for } r_p/\sqrt{2} < r_a < r_p, \tag{32}$$

where $\mathcal{N}_1$ is the normalization term of the binary mass ratio distribution (Eq. 24): $\mathcal{N}_1 = \int_0^1 q^{\gamma+\beta}(1+q^\alpha)^{3/2}\mathrm{d}q$.

For $i = 2$, there is no analytic solution, because evaluating the integral requires imposing the constraint that $r_a = r_p(1+q^{-\alpha})^{-1/2}q^{-1}$. This equation can be re-written

$$\left(\frac{r_p}{r_a}\right)^2 = q^2 + q^{-\alpha+2}, \tag{33}$$

which has no analytic solution except for special values of α, the mass-luminosity exponent. For $\alpha = 3.5$, our nominal case, semianalytic solutions can be found.

Since our main interest is in understanding the qualitative behavior of the solutions, we focus on a few analytic limiting cases, and then proceed numerically.

Correction to inferred rate density—Recall that the rate density correction factor, X_Γ, is the ratio of the apparent to true rate densities. We consider a "nominal model" in which the stellar population is similar to Sun-like stars in the local neighborhood: BF = 0.44, $\alpha = 3.5$, $\beta = 0$. Our default assumption is also that the occurrence of planets is independent of stellar mass ($\gamma = 0$), so secondaries have the same occurrence rate as primaries and single stars. Under these assumptions, the planetary rate density is

$$\Gamma(r) \approx \delta(r_p)\left(\Lambda_0 + \Lambda_1 + \Lambda_2\right)/3, \tag{34}$$

where the coefficients of 1/3 are accurate to within one percent of the true coefficients. Ignoring binarity, the observer finds an apparent rate density

$$\Gamma_a(r) = c_0\Lambda_0\delta(r_p) + c_1\Lambda_1 I_1(r_a) + c_2\Lambda_2 I_2(r_a), \tag{35}$$

for $c_0 \approx 0.49$, $c_1 \approx 0.32$, $c_2 \approx 0.03$. Evaluating the correction term at $r = r_p$, since $\lim_{q\to 0} p_i(r_a) = 0$ for $i \in \{1, 2\}$,

$$X_\Gamma(r = r_p) \approx \frac{3c_0\Lambda_0}{\Lambda_0 + \Lambda_1 + \Lambda_2}. \tag{36}$$

If all the rates are equal, $X_\Gamma(r = r_p) \approx 0.49$. If there are no planets around the secondaries, $X_\Gamma(r = r_p) \approx 0.74$.

We also run this model in our numerical simulation. The results are shown in Figs. 3 and 4. They produce the same correction factors as predicted analytically. They also show a peak in the inferred rate at $r_p/\sqrt{2}$ from secondaries. This effect requires the observer to incorrectly estimate the host star's radius; if they somehow knew the host star's radius, but did not correct for binary dilution, the peak would instead be at an apparent radius of 0.

To summarize, dilution produces a spectrum of apparent planetary radii. In this model, this produces overestimated rates everywhere except where there are actually planets, where the rate is underestimated by a factor of 2.

3.3. *Model #3: Fixed primaries, varying planets and secondaries*

In this model, as in the previous one, all single and primary stars have identical properties. Only the secondaries have masses, radii, and luminosities that vary between systems: $M \propto R \propto L^{1/\alpha}$. The radii of planets are assigned independently of any host system property, and are sampled from an intrinsic radius distribution, which we take as

$$p_r(r) \propto \begin{cases} r^\delta & \text{for } r \geq 2r_\oplus \\ \text{constant} & \text{for } r \leq 2r_\oplus. \end{cases} \tag{37}$$

Following Howard et al. (2012)'s measurement, we take $\delta = -2.92$. Our "nominal model" remains the same: the binary fraction is 0.44, $\alpha = \beta = \gamma = 0$. We take Λ_i, the occurrence rate integrated over all possible phase space for the i^{th} system type, to be equal for singles, primaries, and secondaries.

For this model, we forgo analytic development and focus only on numerics (Sec. 2.1). The nominal results are shown in Fig. 5. For the assumed planetary and stellar distributions, the inferred rate is underestimated over all radii.

Hot Jupiter Occurrence Rates—Taking Fig. 5 and counting the number of planets per star with $r > 8r_\oplus$, we can compare the true and inferred hot Jupiter occurrence rates. Under the above assumptions, the true rate is 9.1 hot Jupiters per thousand stars. The inferred rate is 6.9 per thousand stars. This means that the inferred rate underestimates the true rate by a multiplicative factor of $\sim$1.3.

However, this result only applies under the assumption that $\Lambda_0 = \Lambda_1 = \Lambda_2$. If hot Jupiters are less common around lower mass stars, it would be more sensible

to consider $\Lambda_2 < \Lambda_0$, while letting single stars and primaries host planets at the same rate. Therefore in Fig. 6 we let Λ_2 vary, and show the resulting inferred and true hot Jupiter ($r > 8r_\oplus$) rates. The result is that the inferred rate is nearly independent of Λ_2 – this is because most ($< 1/10$) secondaries are not searchable, and so their completeness fraction is much smaller than that of primaries or single stars. This means far fewer detected hot Jupiters orbit secondaries, and so they hardly affect the inferred rate. While the "true rate" across the entire population is highly dependent on Λ_2, as one would expect from Eq. 5, the rate around singles and primaries (green line) is independent of that around secondaries. Assuming that RV surveys are measuring the true rate around single stars (or primaries), this suggests that binarity might contributes to the HJ rate discrepancy at the $\sim 0.2\%$ level, independent of the HJ rate around secondaries.

The Rate of Earth Analogs—The rate (density) of Earth-like planets orbiting Sun-like stars has been independently measured by Youdin, Petigura, Dong & Zhu, Foreman-Mackey et al., and Burke et al., (2011, 2013, 2013, 2014, and 2015 respectively). These efforts have found that the one-year terrestrial planet occurrence rate varies between ≈ 0.03 and ≈ 1 per Sun-like star, depending on assumptions that are made when retrieving the rate (Burke et al. 2015's Fig. 17).

Our model does not explicitly include the rate density's period-dependence, because stellar binarity does not appreciably bias the period-dependence of occurrence rates measured by transit surveys[6]. Instead, it allows us to evaluate the difference in the apparent and true rate as a function of radius (Fig. 5). At Earth's radius, the result is that the inferred rate is $0.84\times$ the true rate around single stars, assuming that the Λ_i's are equal. Similar to the above case of the hot Jupiters, if we vary the true Λ_2 while keeping $\Lambda_0 = \Lambda_1$, the ratio of the inferred to true rate around single stars hardly changes; $(\Lambda_{\rm inferred}/\Lambda_1)_{r=r_\oplus} \approx 0.135$, to within a few percent, independent of Λ_2 (shown in Fig. 7). The ratio of the inferred to the true rate, $(\Lambda_{\rm inferred}/\Lambda)_{r=r_\oplus}$ varies substantially, but by at most 50% in the (unrealistic) limiting case that secondaries do not host planets.

4. DISCUSSION

Does a detected planet orbit the primary or secondary?—Ciardi et al. (2015) studied the effects of stellar multiplicity on the planet radii derived from transit surveys. They modeled the problem for *Kepler* objects of interest by matching a population of binary and tertiary companions to KOI stars, under the assumption that the KIC-listed stars were the primaries. They then computed planet radius correction factors assuming that *Kepler*-detected planets orbited the primary or companion stars with

[6] Note that stellar binarity does bias the *intrinsic* planet occurrence as a function of planetary and binary periods. This is expected from dynamical stability limits in $\geq$3 body systems, and has been observationally confirmed (theory by Holman & Wiegert 1999, and others including Gongjie; confirmation from Wang et al 2014a, 2014b, Kraus et al 2014). However our statement is that inferred rates as a function of planet period should be negligibly affected by this, given the geometric bias against long-period transit detections, and the fact that the period distribution of solar-type binaries peaks at ≈ 100 years (Raghavan et al 2010, Fig. 13).

equal probability (their Sec. 5). Under these assumptions, they found that any given planet's radius is on average underestimated by a multiplicative factor of 1.5.

Our models show that assuming a detected planet has equal probability of orbiting the primary or secondary leads to an overstatement of binarity's population-level effects. A planet orbiting the secondary does lead to extreme corrections, but these cases are rare outliers, because the searchable volume for secondaries is so much smaller than that for primaries. Phrased in terms of the completeness, in our Model #3 only $\sim 6\%$ of selected secondaries are searchable, compared to $\sim 60\%$ of selected primaries. This means that when high-resolution imaging discovers a binary companion in system that hosts a detected transiting planet, the planet is much more likely to orbit the primary. This statement is independent of the fact that planets are often confirmed to orbit the primary by inferring the stellar density from the transit duration.

On the utility for future occurrence rate measurements—Though they will be difficult to distinguish from false positives, *TESS* is expected to discover over 10^4 giant planets (Sullivan et al. 2015). One possible use of this overwhelmingly large sample will to measure an occurrence rate of short-period giant planets. Our work indicates that if this measurement is to be more precise than $\sim 30\%$, binarity cannot be neglected.

What about Kepler?—Barclay & Collaborator (in preparation) have performed the exercise of taking stars selected by the *Kepler* team, pairing them with a population of secondaries, injecting a realistic distribution of planet radii, and then comparing the inferred occurrence rates with the true ones. In their model, they find that binarity leads to an inferred rate of Earth-sized planets $\approx 10\%$ less than the true rate. In our Model #3, if all Λ_i's are equal (a plausible assumption in the lack of evidence to the contrary), the underestimate is by 16%.

5. CONCLUSION

This study presented three simple models for the effects of binarity on occurrence rates measured by transit surveys. The most realistic of these models (Model #3) suggests that binarity leads to underestimates of occurrence rates at less than 30% relative error. The model indicates that hot Jupiter rates measured by transit surveys are biased to infer ≈ 2 fewer hot Jupiters per thousand single stars than surveys that only measure occurrence rates about single stars (*i.e.,* radial velocity surveys). Our third model also indicates that binarity's effects on the measured occurrence rates of Earth-sized planets are far smaller than current systematic uncertainties. Though our models are simplistic, their agreement with Barclay & Collaborator's recent detailed simulations indicate that they capture the essential ingredients.

It was a pleasure to share discussions with Kento Masuda, who pointed us in this direction, and helped clarify that something like this was worth studying.

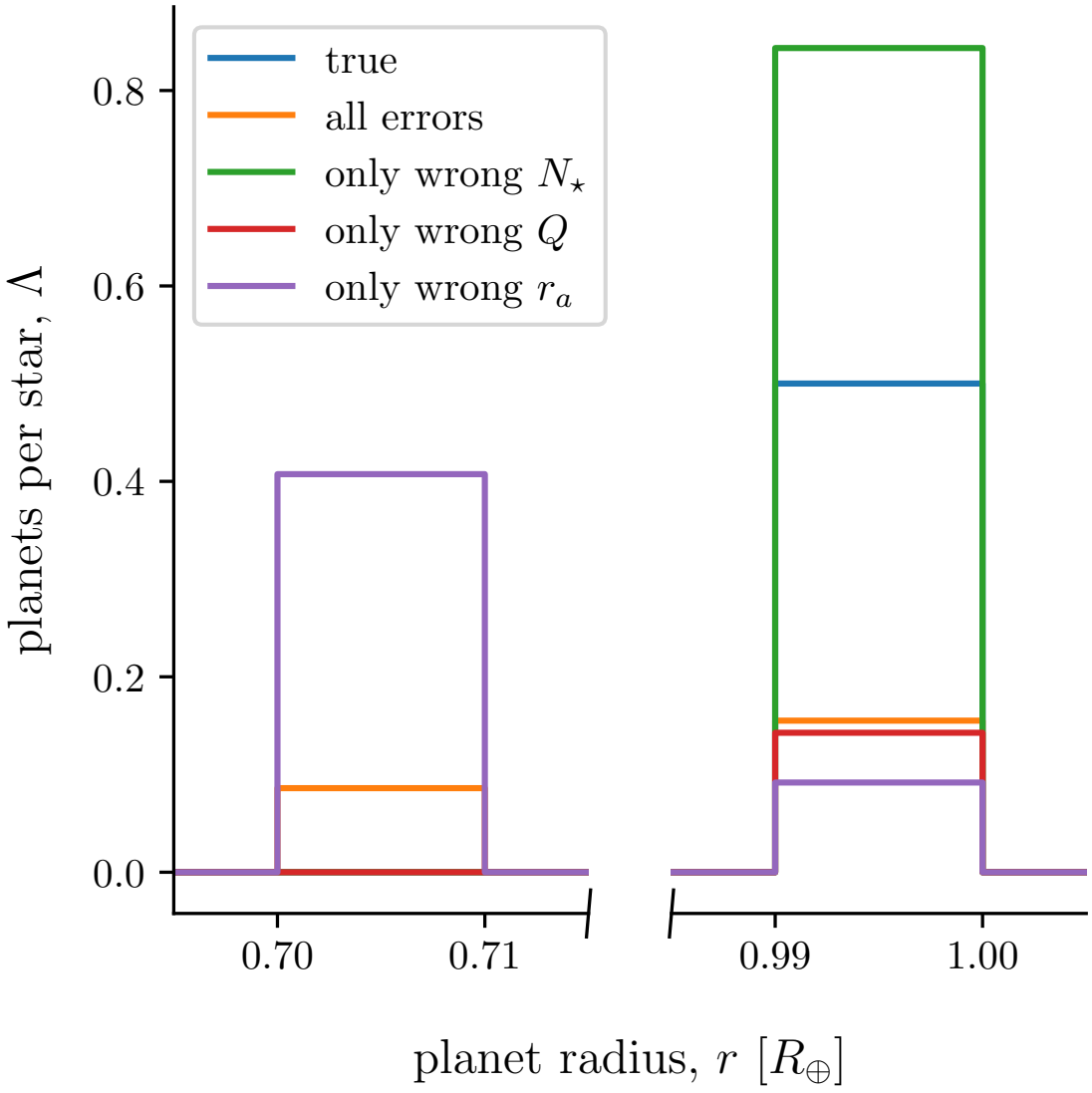

Figure 1. Inferred planet occurrence rates as a function of planet radius in Model #1. This model has fixed stars, fixed planets, and twin binaries. If the true planet radius is r_p, all planets detected in binaries will have apparent radii $r_a = r_p/\sqrt{2}$. Only 1 in 8 selected binaries is actually searchable (see Sec. 3.1). To help illustrate the individual effects of errors, we separated them: "only wrong $N_\star$" means the only error is an incorrectly assumed number of selected stars; "only wrong Q" means the only error is an incorrectly assumed detection efficiency (including both miscalculated transit probabilities and fraction of selected stars that are searchable); "only wrong r_a" means the only error is misinterpretation of planetary radii, due to both transit depth dilution and also wrongly assumed host star radii.

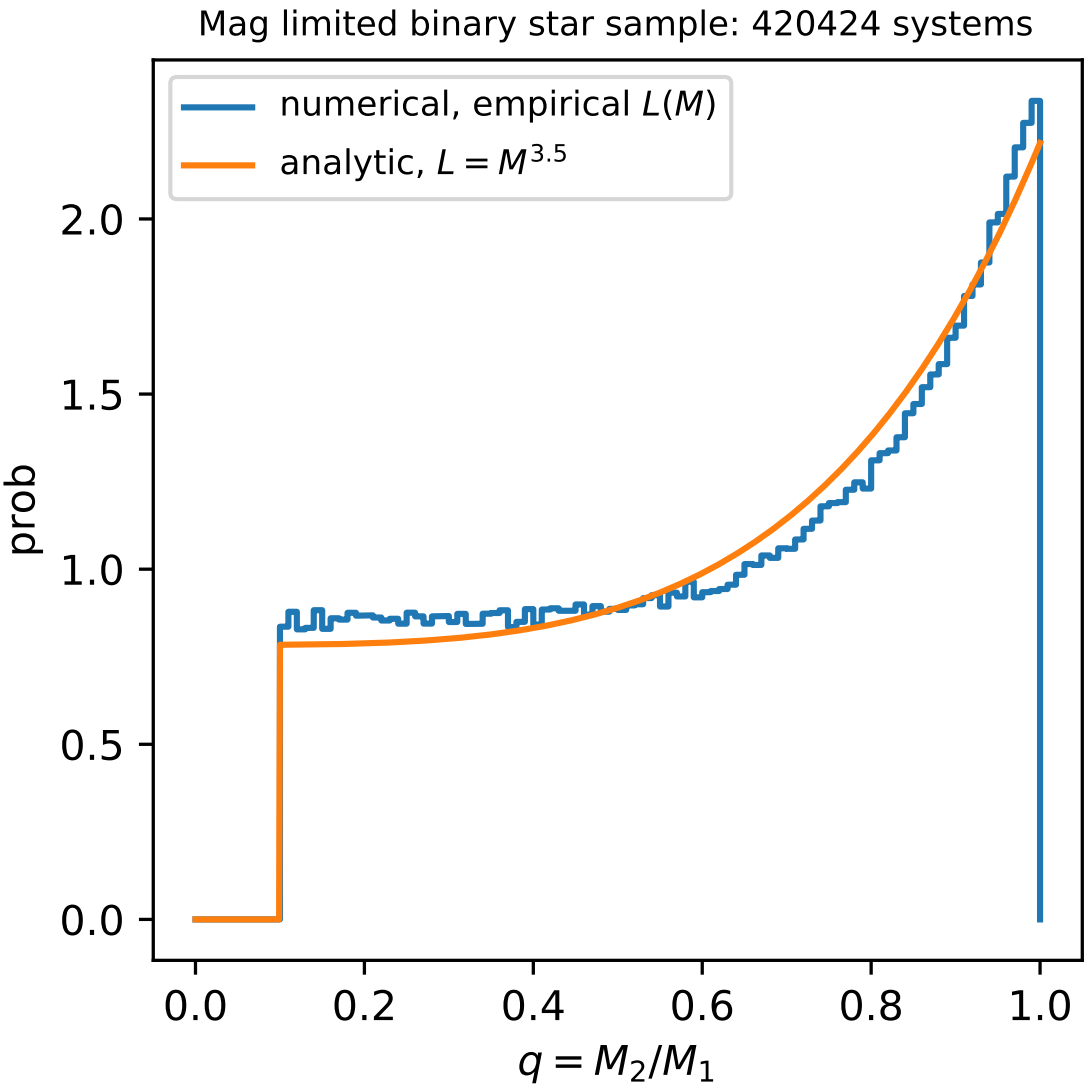

Figure 2. The distribution of the mass ratio for a magnitude limited sample of binary stars. The underlying mass ratios are drawn from a uniform distribution in a volume-limited sample, quite similar to that of Raghavan et al. (2010)'s Fig. 13. The entire bias can be understood analytically (Eq. 24, with $\gamma = 0$). The numerical comparison uses a realistic empirical mass-luminosity relation rather than $L = M^{3.5}$, found by fitting main sequence dwarf mass luminosity data compiled from Torres et al. [2009] and Benedict et al. [2016].

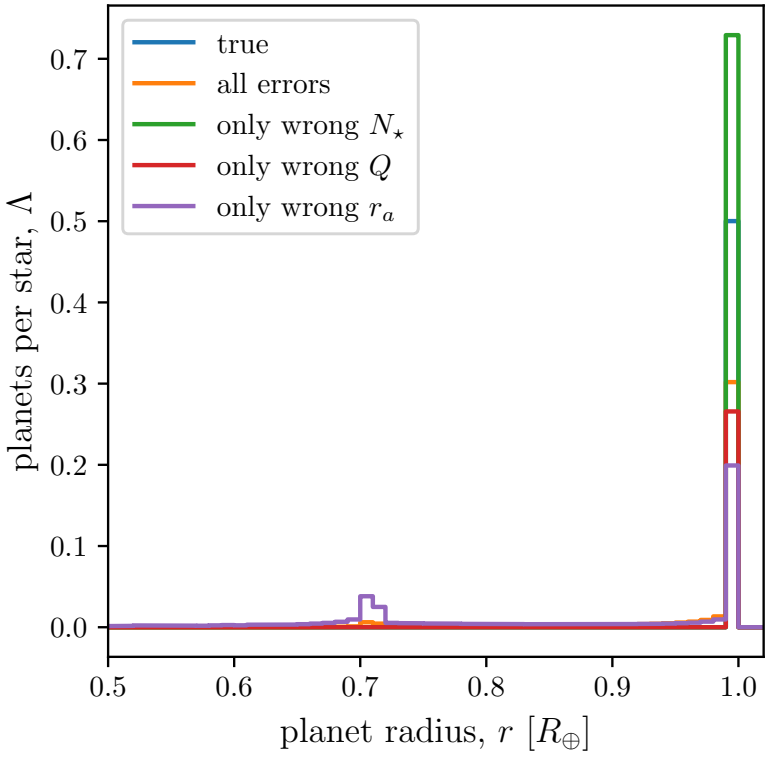

Figure 3. Inferred planet occurrence rates as a function of planet radius in Model #2. This model has fixed planets and primaries, but varying secondaries.

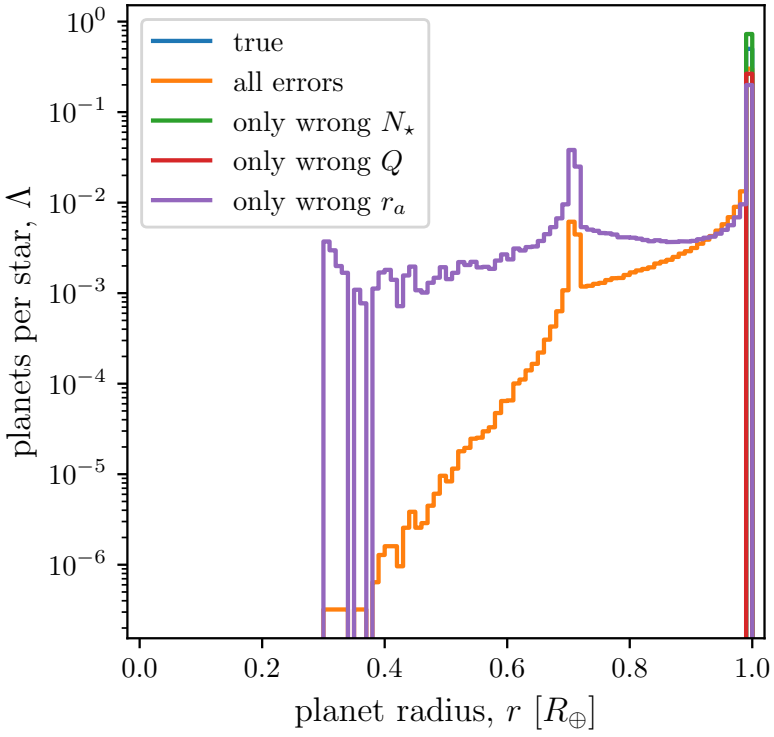

Figure 4. Same as Fig. 3, but with logarithmic y-axis, and different x scale.

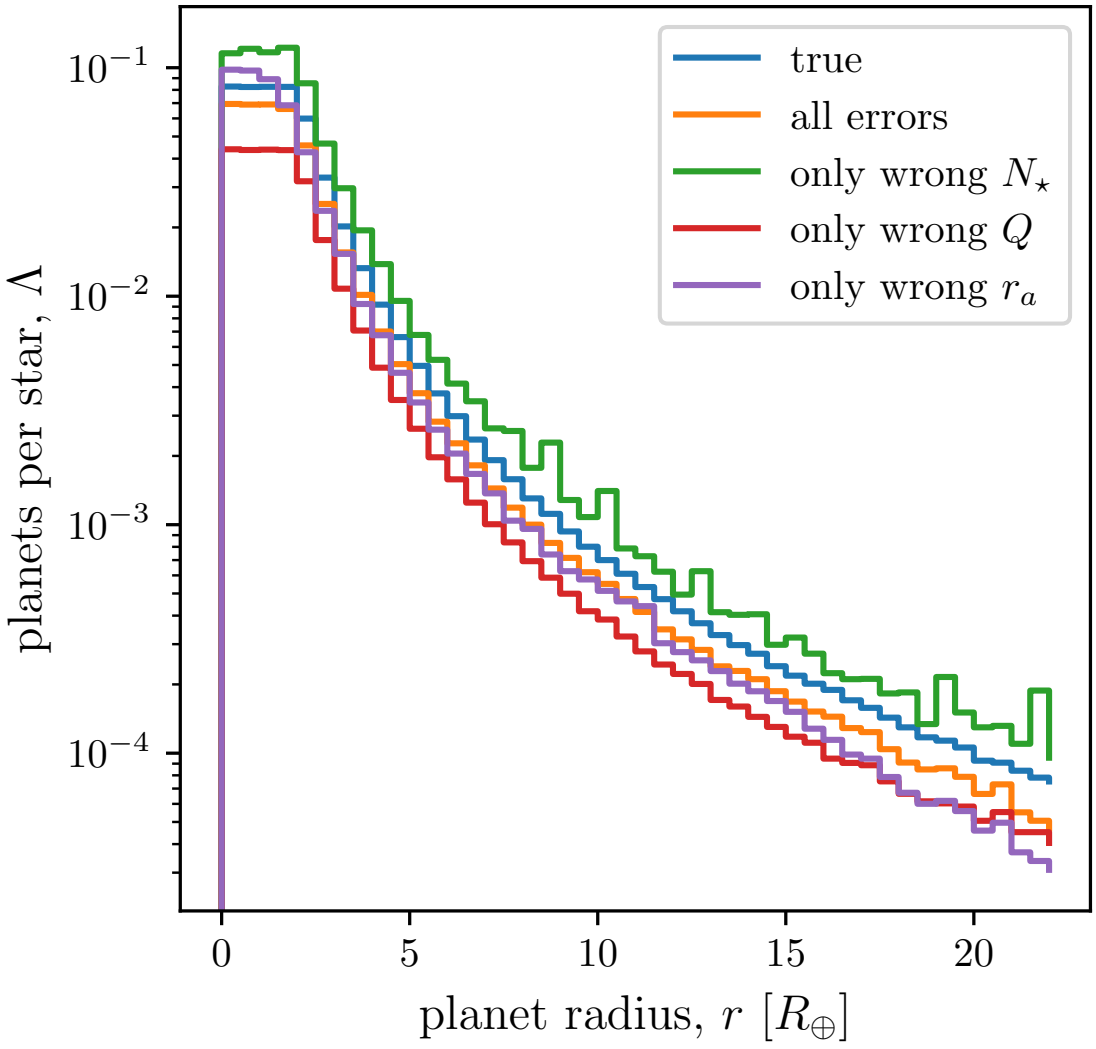

Figure 5. Inferred planet occurrence rates as a function of planet radius in Model #3. This model has fixed primaries and single stars, but varying secondaries. The true planet radius distribution is a power law with exponent -2.92 above $2R_\oplus$, below which it is uniform (e.g., Howard et al., 2012).

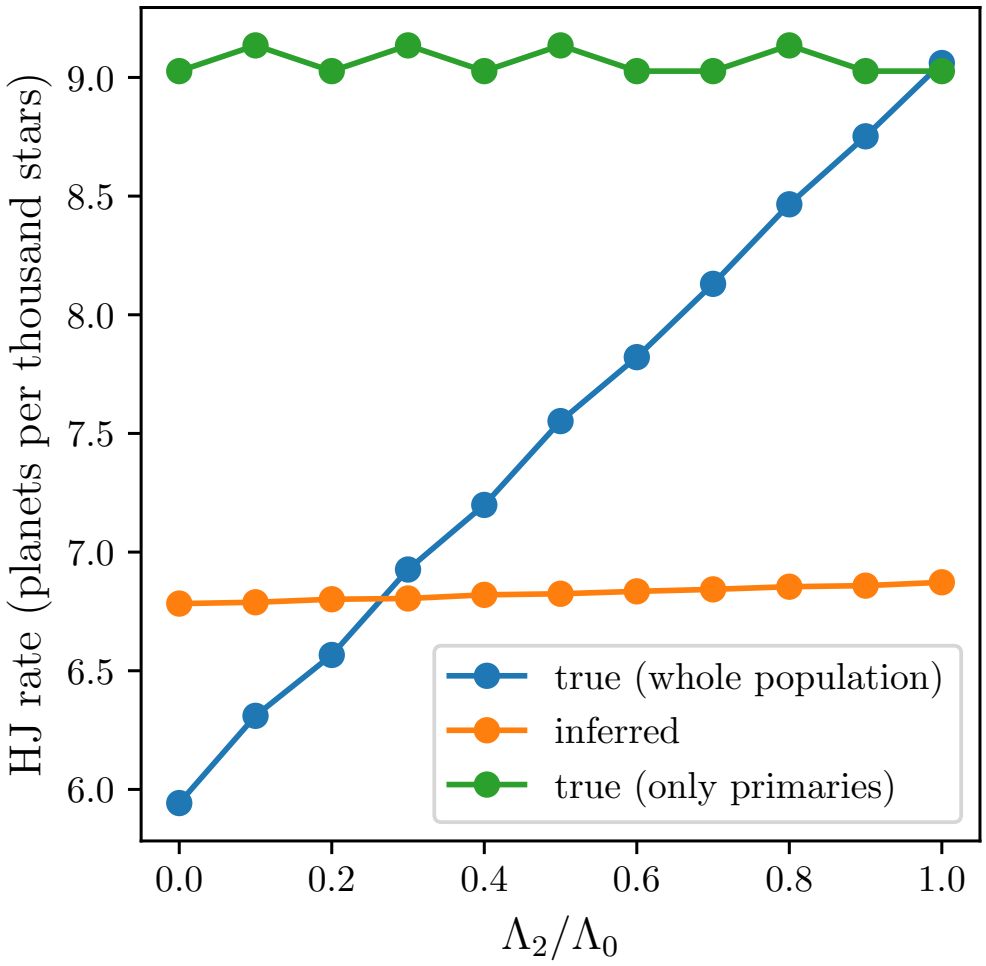

Figure 6. Λ_i is the occurrence rate integrated over all possible phase space for the i^{th} system type. $\Lambda_2/\Lambda_0 = 1$ corresponds to an equal number of planets per secondary as per single star; $\Lambda_2/\Lambda_0 = 0$ corresponds to secondaries not having any planets. In our Model #3, though the true HJ occurrence rate (Eq. 5) is highly dependent on Λ_2, the inferred rate hardly depends on whether secondaries have HJs. This means that the "correction factor" between the inferred rate and the true rate around single stars is underestimated by a multiplicative factor of ≈ 1.3, independent of the HJ rate around secondaries. The "HJ rate" is the summed rate from Fig. 5 above $8r_\oplus$.

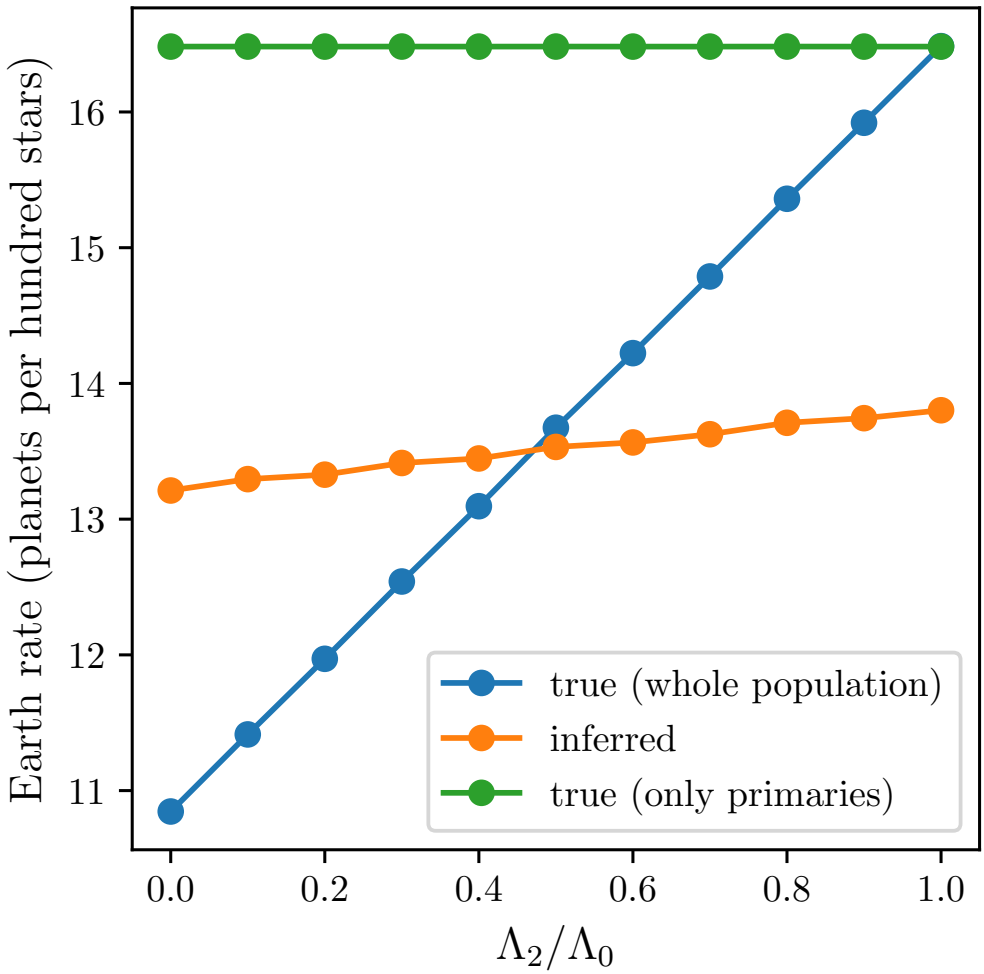

Figure 7. Same as Fig. 6, but for Earth-sized planets. The absolute values given on the y-axis found by summing the rate from Fig. 5 for planetary radii from 0.5 to $1.5r_\oplus$ (this is a toy model, and they do not reflect an actual determination of $\eta_\oplus$). The relative values show that the inferred rate for Earths is roughly independent of the occurrence rate (integrated over all radii) around secondaries. However, it is systematically lower than the true rate around single and primary stars, by $\approx 20\%$.

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+ 2005	12±1	$a < 0.1\,\mathrm{AU}; P \lesssim 10\,\mathrm{day}$
Cumming+ 2008	15±6	–
Mayor+ 2011	8.9±3.6	–
Wright+ 2012	12.0±3.8	–
Gould+ 2006	$3.1^{+4.3}_{-1.8}$	$P < 5\,\mathrm{day}$
Bayliss+ 2011	10^{+27}_{-8}	$P < 10\,\mathrm{day}$
Howard+ 2012	4±1	$P < 10\,\mathrm{day}; r_p = 8 - 32 r_\oplus$; solar subset[a]
–	5±1	solar subset extended to $Kp < 16$
–	7.6±1.3	solar subset extended to $r_p > 5.6 r_\oplus$.
Petigura+ 2017	XX±YY	?

NOTE— The upper four results are from radial velocity surveys; the latter six are from transit surveys. Many of these surveys selected different stellar samples.

[a] Howard+ 2012's "solar subset" was defined as *Kepler*-observed stars with $4100\,\mathrm{K} < T_\mathrm{eff} < 6100\,\mathrm{K}$, $Kp < 15$, $4.0 < \log g < 4.9$. Their rate selected planets with measured signal to noise > 10.