

Supplement for Potential Test-Negative Case-Control Study Bias in Outbreak Settings: Application to Ebola vaccination in DRC

Carl A. B. Pearson^{1,2,*}

W. John Edmunds¹

Thomas J. Hladish³

Rosalind M. Eggo¹

¹Department of Infectious Disease Epidemiology &
Centre for Mathematical Modelling of Infectious Diseases
London School of Hygiene & Tropical Medicine

²South African Centre for Epidemiological Modelling and Analysis
Stellenbosch University

³Department of Biology & Emerging Pathogens Institute
University of Florida

*correspondence to: carl.pearson {at} lshtm.ac.uk

6 January, 2020

Contents

S1 Overview	3
S2 Generalisation of Nomenclature	3
S2.1 Model Diagram	4
S2.2 Population Heterogeneity	5
S2.3 Exposure, Protection, Infection, and Recruitment	5
S2.4 Recruitment-Related Parameters	6
S2.5 Summary Recruitment Categories	6
S2.6 Expanded TNCC Intervention Efficacy Estimator	6
S3 Efficacy Estimator Bias	7
S3.1 Test-Positive Odds	7
S3.2 Test-Negative Odds	9
S3.3 Relaxing Assumption that Prevented Infections Remain Secondary Recruits	11
S4 Total Estimator Bias & Limiting Scenarios	12
S4.1 True efficacy, E , limits	13
S4.2 Participating fraction, p_{in} , limits	13
S4.3 Secondary Case Recruitment, p_t , limits	15
S4.4 Lower Limit on Secondary Relative Recruiting, ρ	15
S5 Translation of Limits to Recruitment Constraints	15
S5.1 Attempting to Limit Recruitment to Participants Only	15
S5.2 Attempting to Limit to Primary Recruitment Only	25
S6 Calculation of Coverage, L, and Participation, p_{in}	36
S6.1 Measures During Intervention Distribution	36
S7 Alternative Scenario Translation	38

1 S1 Overview

2 A Test-Negative Case Control (TNCC) study has been proposed to evaluate a new EVD vaccine during the
3 ongoing EVD epidemic in Eastern Democratic Republic of Congo (DRC). The main text discussed a model
4 of such a study: we described a population of individuals i) who heterogeneously receive a study vaccine, ii)
5 some of whom subsequently exhibit symptoms of Ebola Virus Disease (EVD) or have contact with a known
6 EVD case, and thus iii) are identified by either self-reporting and contact-tracing processes as part of an
7 outbreak response, and finally iv) are tested for EVD, making them recruitable for a TNCC study of that
8 vaccine. Particularly, we explored the potential for bias due to route of recruitment into the study, and
9 due to heterogeneity in vaccine distribution. As we noted in the discussion, TNCC studies could be used to
10 evaluate other kinds interventions.

11 To support application of the model to other contexts, we use more general terms in this Supplement
12 (Section S2). Using those terms, we provide derivations of equations quoted in the main text (Sections
13 S3-S5), additional results (Section S6), and translate the model to another example (Section S7).

14 S2 Generalisation of Nomenclature

15 We consider a generic *study intervention*. The study intervention is in addition to any other outbreak control
16 measures that might be ongoing. As in the main text, this study intervention is heterogeneously distributed
17 at an individual level, which we represent with participation status and intervention coverage. The TNCC
18 study goal is to determine the intervention efficacy.

19 The main text discusses self-reporting and contact-tracing as particular recruiting routes for EVD, which
20 more generally are a random *primary* process and a reactive *secondary* process, respectively; we also use these
21 qualifiers to distinguish the associated exposure processes to the pathogen of interest. For both routes, we
22 still assume a highly sensitive and specific test for identifying infections with the target pathogen. For EVD,
23 recruitment is symptom-related, but it does not have to be for all pathogens, so here we discuss intervention
24 efficacy in terms of infection rather than disease.

25 The key assumptions in this model are:

- 26 • exposure to the target pathogen is identical for all individuals
- 27 • if infected by exposure, the probability of detection by the primary process is identical for all individuals
- 28 • the rate of exposure to any other pathogens that could result in testing (and therefore recruitment) is
29 the same for all individuals
- 30 • all secondary recruits associated with an initial primary case have the same participation status as
31 that primary case
- 32 • the secondary transmission and recruitment process is identical for all individuals

33 S2.1 Model Diagram

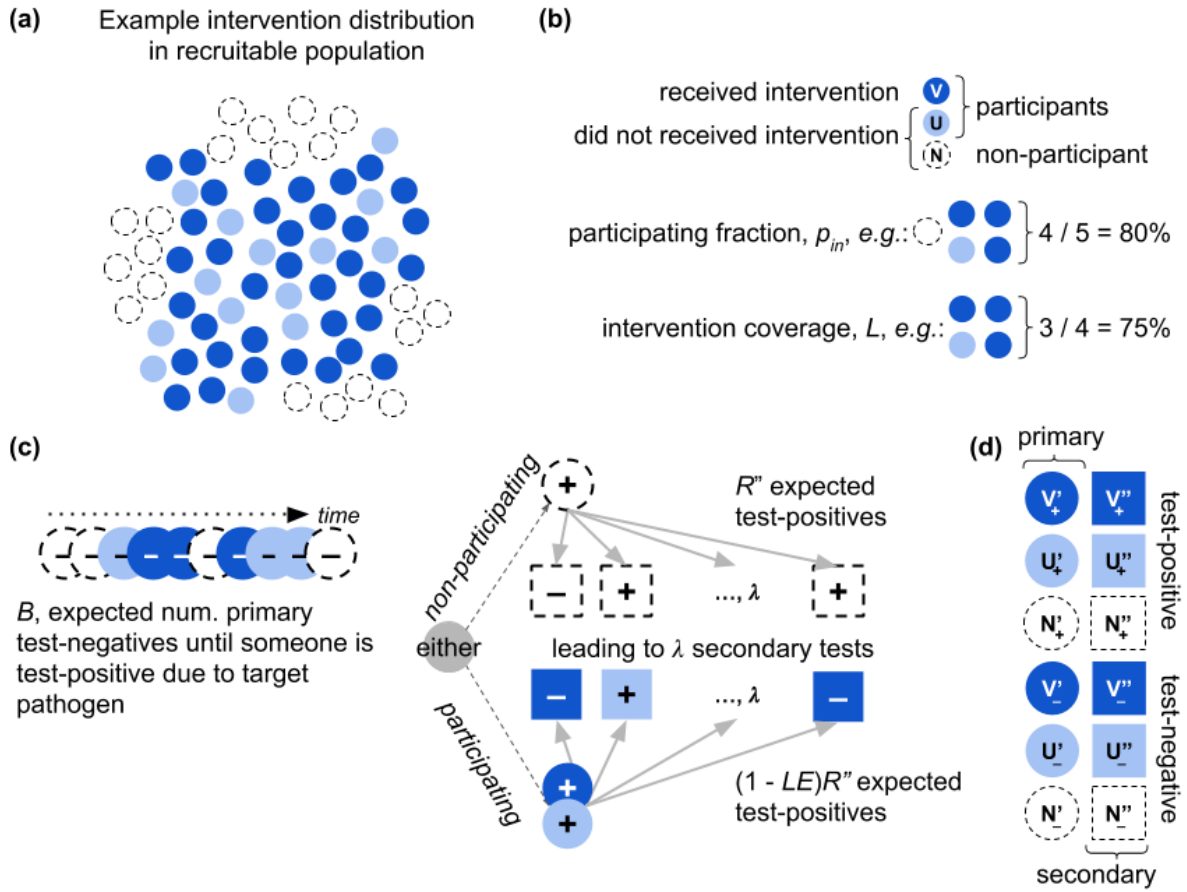


Figure S1: **Model Diagram.** This diagram illustrates the conceptual model of the population and recruitment process, showing: how the intervention might be distributed in the population (panel (a)), with the corresponding intervention term definitions for *participating fraction*, p_{in} , and *intervention coverage*, L (panel (b)); and how the primary process proceeds by finding an expected number of test-negatives, B , until finding a test-positive case, which initiates the secondary process leading to an expected number of additional tests, λ , of which R'' are expected to be test-positive in the absence of the study intervention (panel (c)), with the corresponding to summary categories of different routes to recruitment into the study (panel (d)).

34 S2.2 Population Heterogeneity

35 There are two types of individuals in the population: *non-participating* and *participating*. Non-participating
36 individuals do not receive the study intervention. Participating individuals randomly receive the intervention
37 with some probability. Aside from participation and vaccination status, individuals are identical: they have
38 the same exposure risk for the target pathogen, same primary testing rate given exposure and non-exposure,
39 and same secondary distribution and probability of exposing those individuals.

40 In the equations that follow, we label counts of individuals corresponding to their participation and inter-
41 vention status. The totals of these categories determine the population heterogeneity characteristics. Total
42 numbers are:

43 V , the number of participating individuals that received the intervention

44 U , the number of participating individuals that did not receive the intervention

45 C , of the recruitable population, those that are participating; $C = V + U$

46 N , individuals that did not participate

47 T , total potentially recruitable study population; $T = C + N$

48 p_{in} , the participating fraction for the study, $p_{\text{in}} = \frac{C}{T}$; used as participation probability for individuals

49 L , the intervention coverage in the participating population, $L = \frac{V}{C}$; used as the probability participating
50 individuals have the intervention

51 S2.3 Exposure, Protection, Infection, and Recruitment

52 During the outbreak, individuals $\{V, U, N\}$ are potentially exposed and tested, and thus recruited in the
53 study, via two routes: i) the primary route, where individuals are exposed and tested randomly, independent
54 of any association with an identified case; ii) the secondary route, where individuals are exposed by and
55 tested because of connection with an identified case.

56 Test-positive individuals found by the primary process and those they associate with as detect by the
57 secondary process have the same participation status in the model. Therefore primary test-positives from
58 the participating population only interact via the secondary process with other participating individuals,
59 and likewise for non-participating primary test-positives.

60 Relative to transmission of the target pathogen, all infections discovered by secondary route are assumed to
61 result from exposures due to the associated primary case. Primary cases are assumed to result from a random
62 exposure process (*i.e.* exposing individuals of the different types according to the relative proportions in the
63 population).

64 All exposures result in infections, unless protected by the study intervention; potential infections prevented
65 by other outbreak response measures (which benefit all individuals in the study population equally) are
66 assumed to not be exposures in the context of the model. For both primary and secondary exposures, an
67 individual that has received the intervention may avoid infection with probability corresponding to the study
68 intervention efficacy. Exposed individuals who have not received the intervention become infected.

69 The intervention is also assumed to have no impact on infections that are not by the target pathogen, even
70 where those infections could lead to a test. Finally, neither process is assumed to observe all of the relevant

71 events (*i.e.*, those that lead to primary or secondary recruitment) that occur, just that observation rates do
 72 not differ by participation status.

73 S2.4 Recruitment-Related Parameters

74 These real-world infection and testing processes are stochastic, and we consider the asymptotic expected
 75 (*i.e.*, average) outcomes of these processes. We define the following set of parameters corresponding to
 76 expected counts of recruitment-related events per primary test-positive, and the probability of preventing
 77 infection.

78 $B + 1$: the expected number of tests from the primary process required to find a single test-positive case; *i.e.*,
 79 on average, for $B + 1$ individuals tested via primary route, B are test-negative and 1 is test-positive.

80 λ : the expected number of tests from the secondary process, for each case identified by the primary process.

81 R'' : the expected number of test-positives in the contacts of identified cases when there is no intervention.

82 E : the intervention efficacy; *i.e.*, the probability that an individual who has received the intervention will
 83 avoid infection when exposed (relative to an individual without the intervention) either via the primary
 84 or secondary routes

85 S2.5 Summary Recruitment Categories

86 In addition to the participation and intervention states, we denote individuals by their routes to testing
 87 and test outcomes, with super- and subscripts respectively. These are summarized earlier in Fig. S1, and as
 88 follows:

89 $\frac{T'_+, N'_+}{C'_+=V'_+U'_+}$: among corresponding individuals: testing positive, recruited via primary process

90 $\frac{T''_+, N''_+}{C''_+=V''_+U''_+}$: from corresponding populations, testing positive, recruited via secondary process

91 $\frac{T'_-, N'_-}{C'_-=V'_-U'_-}$: from corresponding populations, testing negative, recruited via primary process

92 $\frac{T''_-, N''_-}{C''_-=V''_-U''_-}$: from corresponding populations, testing negative, recruited via secondary process

93 S2.6 Expanded TNCC Intervention Efficacy Estimator

94 The conventional TNCC efficacy estimator stratifies the recruited population into four groups, made from
 95 two binary distinctions: received the intervention versus not, and test-positive for the pathogen of interest
 96 versus test-negative. That efficacy estimator is:

$$\text{estimated efficacy} = 1 - \frac{\# \text{ intervention, test-positive}}{\# \text{ non-intervention, test-positive}} \times \frac{\# \text{ non-intervention, test-negative}}{\# \text{ intervention, test-negative}}$$

97 In terms of the types of individuals defined at the end of Section S2.3, this equation becomes:

$$\hat{E} = 1 - \frac{V'_+ + V''_+}{N'_+ + N''_+ + U'_+ + U''_+} \times \frac{N'_- + U'_- + N''_- + U''_-}{V'_- + V''_-} \quad (\text{S1})$$

100 which can be thought of as two odds of having received the intervention, each conditional on the two possible
 101 test outcomes. We will refer to these as the test-positive odds and the test-negative odds.

102 S3 Efficacy Estimator Bias

103 To determine the potential bias of \hat{E} for a TNCC study in an outbreak setting, we need to translate Eq. S1
 104 from being in terms of total individual counts, into the expected totals, given the study target intervention
 105 and outbreak conditions. We will examine each of the two odds terms in turn, then combine the results.

106 S3.1 Test-Positive Odds

107 Starting with the test-positive odds term, we first factorise (by C'_+ , or N'_+), such that most terms are expressed
 108 as proportions:

$$\frac{V'_+ + V''_+}{N'_+ + N''_+ + U'_+ + U''_+} = \frac{C'_+ \left(\frac{V'_+}{C'_+} + \frac{V''_+}{C'_+} \right)}{N'_+ \left(1 + \frac{N''_+}{N'_+} \right) + C'_+ \left(\frac{U'_+}{C'_+} + \frac{U''_+}{C'_+} \right)} \quad (\text{S2})$$

109 Recall we defined R'' as the expected number of cases that would be found via the secondary route among
 110 individuals that had not received the study intervention (*i.e.*, non-participating individuals or individuals
 111 outside of the recruitable population). The proportion $\frac{N''_+}{N'_+}$ is total number of secondary cases over the
 112 total number of primary cases (among non-participating individuals), which is also the average number of
 113 secondary cases per primary case, so we can substitute $R'' = \frac{N''_+}{N'_+}$. Since we have restricted the study to a
 114 scenario where only a single primary case exposes secondary cases, there is neither indirect protection or
 115 force of infection from multiple sources. Thus, amongst participating individuals (*i.e.* those in C), R'' will
 116 be reduced on average by the probability that exposed individuals are intervention recipients and protected.
 117 This probability is equal to coverage, L , multiplied by the efficacy, E . Therefore, $\frac{C''_+}{C'_+} = R''(1 - LE)$ in the
 118 participating population.

119 Using these substitutions and introducing some identity multipliers, $1 = \frac{C''_+}{C''_+}$, we can rewrite Eq. S2 as:

$$\frac{C'_+ \left(\frac{V'_+}{C'_+} + \frac{C''_+}{C'_+} \frac{V''_+}{C''_+} \right)}{N'_+ \left(1 + \frac{N''_+}{N'_+} \right) + C'_+ \left(\frac{U'_+}{C'_+} + \frac{C''_+}{C'_+} \frac{U''_+}{C''_+} \right)} = \frac{C'_+ \left(\frac{V'_+}{C'_+} + \frac{V''_+}{C''_+} R''(1 - LE) \right)}{N'_+ (1 + R'') + C'_+ \left(\frac{U'_+}{C'_+} + \frac{U''_+}{C''_+} R''(1 - LE) \right)} \quad (\text{S3})$$

120 Next, we show how the proportions between the counts of the participating individuals can be substituted
 121 to express the odds in terms of model parameters. Starting with $\frac{V'_+}{C'_+}$: this is the probability of an individual
 122 receiving the intervention, conditional on there being an initial infection in the participating population.
 123 Though the exposure probabilities are the same between participating and non-participating populations, if
 124 $LE > 0$, the probability that an exposure results in an infection is lower. Given an exposure event:

$$\begin{aligned} \frac{V'_+}{C'_+} &= P(\text{received intervention} \mid \text{is infected \& participating}) = P(i \in V \mid +, i \in C) \\ &= \frac{P(+ \mid i \in V) \times P(i \in V \mid i \in C)}{P(+ \mid i \in C)} = \frac{(1 - E) \times L}{(1 - L) + L(1 - E)} = \frac{(1 - E)L}{1 - LE} \end{aligned}$$

123 We can use the same logic for the secondary exposures: conditional on a secondary exposure that could
 124 result in infection, the same relationship applies. To solve for $\frac{U'_+}{C'_+}$, we take the complements. Therefore,
 125 these four ratios are:

$$\begin{aligned}\frac{V'_+}{C'_+} &= \frac{V''_+}{C''_+} = \frac{(1-E)L}{1-LE} \\ \frac{U'_+}{C'_+} &= \frac{U''_+}{C''_+} = \frac{1-L}{1-LE}\end{aligned}\tag{S4}$$

126 which means we can further substitute in Eq. S3:

$$\begin{aligned}\frac{C'_+ \left(\frac{V'_+}{C'_+} + \frac{V''_+}{C''_+} R''(1-LE) \right)}{N'_+ (1+R'') + C'_+ \left(\frac{U'_+}{C'_+} + \frac{U''_+}{C''_+} R''(1-LE) \right)} &= \frac{C'_+ \frac{V'_+}{C'_+} (1+R''(1-LE))}{N'_+ (1+R'') + C'_+ \frac{U'_+}{C'_+} (1+R''(1-LE))} \\ &= \frac{C'_+ \frac{(1-E)L}{1-LE} (1+R''(1-LE))}{N'_+ (1+R'') + C'_+ \frac{1-L}{1-LE} (1+R''(1-LE))} \\ &= \frac{\frac{C'_+ (1-E)L}{T'_+ (1-LE)}}{\frac{N'_+ (1+R'')}{T'_+ (1+R''(1-LE))} + \frac{C'_+ (1-L)}{T'_+ (1-LE)}}\end{aligned}\tag{S5}$$

127 Like for determining the fraction of infections in participating individuals that did or did not receive the
 128 intervention (Eq. S3.1), we can also use Bayes Theorem to find the relative fractions of infections that
 129 occurred in individuals that did or did not participate, $\frac{C'_+}{T'_+}$ and $\frac{N'_+}{T'_+}$:

$$\begin{aligned}\frac{C'_+}{T'_+} &= P(i \in C|+) = \frac{P(+|i \in C)P(i \in C)}{P(+)} = \frac{(1-LE)p_{\text{in}}}{(1-p_{\text{in}}) + (1-LE)p_{\text{in}}} = \frac{(1-LE)p_{\text{in}}}{1-LEp_{\text{in}}} \\ \frac{N'_+}{T'_+} &= P(i \in N|+) = 1 - P(i \in C|+) = \frac{1-p_{\text{in}}}{1-LEp_{\text{in}}}\end{aligned}\tag{S6}$$

130 which means that,

$$\begin{aligned}\frac{\frac{C'_+ (1-E)L}{T'_+ (1-LE)}}{\frac{N'_+ (1+R'')}{T'_+ (1+R''(1-LE))} + \frac{C'_+ (1-L)}{T'_+ (1-LE)}} &= \frac{1-LEp_{\text{in}}}{1-LEp_{\text{in}} (1-p_{\text{in}})} \frac{(1-LE)p_{\text{in}} \frac{(1-E)L}{1-LE}}{\frac{(1+R'')}{1+R''(1-LE)} + (1-LE)p_{\text{in}} \frac{1-L}{1-LE}} \\ &= \frac{p_{\text{in}}(1-E)L}{(1-p_{\text{in}}) \frac{(1+R'')}{1+R''(1-LE)} + p_{\text{in}}(1-L)}\end{aligned}\tag{S7}$$

131 and therefore,

$$\frac{V'_+ + V_+}{N'_+ + N_+ + U'_+ + U_+} = \frac{p_{\text{in}}(1-E)L}{(1-p_{\text{in}}) \frac{(1+R'')}{1+R''(1-LE)} + p_{\text{in}}(1-L)}\tag{S8}$$

132 In conventional TNCC studies there is no secondary recruitment. If secondary recruitment were eliminated
 133 under outbreak circumstances, that would imply that $\lambda \rightarrow 0$, which also means that $R'' \rightarrow 0$. During
 134 outbreaks, the secondary process would still occur as part of the response (*i.e.* there would both testing and
 135 case-finding), but the people identified would not be recruited. Under that limit:

$$\begin{aligned} \lim_{R'' \rightarrow 0} \frac{1 + R''}{1 + R''(1 - LE)} &= 1 \\ \lim_{R'' \rightarrow 0} \frac{Lp_{\text{in}}(1 - E)}{(1 - p_{\text{in}}) \frac{1 + R''}{1 + R''(1 - LE)} + p_{\text{in}} - Lp_{\text{in}}} &= \frac{Lp_{\text{in}}(1 - E)}{1 - Lp_{\text{in}}} \end{aligned} \quad (\text{S9})$$

136 which suggests a useful re-arrangement of the final form of the test-positive odds, so it has a clear separation
 137 of the terms which appear in the unbiased estimator (*i.e.* primary recruiting only) and the remaining
 138 factors:

$$\begin{aligned} \frac{Lp_{\text{in}}(1 - E)}{(1 - p_{\text{in}}) \frac{1 + R''}{1 + R''(1 - LE)} + p_{\text{in}}(1 - L)} &= \frac{Lp_{\text{in}}(1 - E)}{(1 - p_{\text{in}}) \frac{1 + R''}{1 + R''(1 - LE)} + p_{\text{in}} - Lp_{\text{in}} + 1 - 1} \\ \text{obtain } 1 - Lp_{\text{in}} \text{ term to factor out...} &= \frac{Lp_{\text{in}}(1 - E)}{(1 - p_{\text{in}}) \frac{1 + R''}{1 + R''(1 - LE)} - (1 - p_{\text{in}}) + (1 - Lp_{\text{in}})} \\ \text{factor other term...} &= \frac{Lp_{\text{in}}(1 - E)}{(1 - Lp_{\text{in}}) + (1 - p_{\text{in}}) \left(\frac{1 + R''}{1 + R''(1 - LE)} - 1 \right)} \\ \text{simplify other term...} &= \frac{Lp_{\text{in}}(1 - E)}{(1 - Lp_{\text{in}}) + (1 - p_{\text{in}}) \left(\frac{LER''}{1 + R''(1 - LE)} \right)} \\ \text{factor out target terms...} &= \frac{Lp_{\text{in}}(1 - E)}{1 - Lp_{\text{in}}} \left[1 + \frac{ER''}{1 + R''(1 - LE)} \frac{L(1 - p_{\text{in}})}{1 - Lp_{\text{in}}} \right]^{-1} \end{aligned} \quad (\text{S10})$$

139 In Eq. S10, we now have only terms that describe the intervention (participation and coverage probabilities,
 140 p_{in} and L , and efficacy E) and epidemiology (R'').

141 S3.2 Test-Negative Odds

142 We assume that the testing criteria for the secondary process is not affected by the presence of the inter-
 143 vention. For example, a contact-tracing-related criteria might be principally about high-risk interactions
 144 rather than particular symptoms, or the symptom threshold might be sufficiently relaxed that almost all
 145 contacts meet it. Similarly, a purely geographical criteria would be unaffected by presence or absence of
 146 the intervention. Thus, in our model all the prevented secondary infections (via intervention efficacy E) are
 147 still recruited by the secondary process as test-negatives. This is a bounding assumption; see the end of this
 148 section for relaxing this assumption.

149 Turning to the test-negative odds, we first replace the primary test-negatives by the contribution from B ,
 150 the average number of test-negatives per test-positive via the primary route. Given that definition, the total
 151 number of primary test-negatives is $T'_- = BT'_+$. Because the intervention has no effect on the causes that lead
 152 to testing negative via the primary route, the representation of individuals follows their proportions in the
 153 population:

$$\frac{N'_- + U'_- + N''_- + U''_-}{V'_- + V''_-} = \frac{BT'_+(1 - Lp_{\text{in}}) + N''_- + U''_-}{BT'_+Lp_{\text{in}} + V''_-} \quad (\text{S11})$$

154 As with the test-positives odds, we can factorise and introduce identity multiples to re-arrange into terms
 155 that we can then use Bayes Theorem to replace with model parameters:

$$\begin{aligned}
\frac{BT'_+(1 - Lp_{\text{in}}) + N''_- + U''_-}{BT'_+Lp_{\text{in}} + V''_-} &= \frac{B(1 - Lp_{\text{in}}) + \frac{1}{T'_+} (N''_- + U''_-)}{BLp_{\text{in}} + \frac{V''_-}{T'_+}} \\
&= \frac{B(1 - Lp_{\text{in}}) + \left(\frac{N'_+}{T'_+} \frac{N''_-}{N'_+} + \frac{C'_+}{T'_+} \frac{U''_-}{C'_+} \right)}{BLp_{\text{in}} + \frac{C'_+}{T'_+} \frac{V''_-}{C'_+}} \tag{S12}
\end{aligned}$$

156 We can use the participating and non-participating fractions of primary test-positives, $\frac{C'_+}{T'_+}$ and $\frac{N'_+}{T'_+}$, from
157 refactoring the test-positive odds (Eq. S6).

158 Amongst non-participating individuals, on average $\lambda - R''$ of recruits from the secondary route will be test-
159 negative. This means that $\frac{N''_-}{N'_+} = \lambda - R''$.

160 This definition also implies that the exposed proportion is $p_t = \frac{R''}{\lambda}$ because R'' individuals are infected per λ
161 secondary individuals. The complementary non-exposed proportion is therefore $1 - p_t = \frac{\lambda - R''}{\lambda}$. This value is
162 like a transmission probability, though that interpretation should be used with caution: the denominator is
163 determined by the secondary observation process, and thus the proportion may not clearly translate to the
164 biological process probability.

Also by definition, amongst participating individuals, only $1 - LE$ of the exposed individuals are infected,
therefore:

$$\frac{C''_-}{C'_+} = (1 - p_t(1 - LE))\lambda = \lambda - R''(1 - LE)$$

165 We again use Bayes Theorem to translate these ratios into model parameter expressions.

$$\begin{aligned}
\frac{U''_-}{C''_-} &= P(\text{is unvaccinated} \mid \text{is not infected \& participating}) = P(i \in U \mid -, i \in C) = \frac{P(-i \in U)P(i \in U \mid i \in C)}{P(-i \in C)} \\
&= \frac{\frac{\lambda - R''}{\lambda}(1 - L)}{\frac{\lambda - R''}{\lambda}(1 - L) + L\left(\frac{\lambda - R''}{\lambda} + \frac{R''}{\lambda}E\right)} = \frac{(\lambda - R'')(1 - L)}{\lambda - R''(1 - LE)} \\
\frac{V''_-}{C''_-} &= 1 - P(i \in U \mid -, i \in C) = \frac{(\lambda - (1 - E)R'')L}{\lambda - (1 - LE)R''} \\
\frac{U''_-}{C'_+} &= \frac{U''_-}{C''_-} \frac{C''_-}{C'_+} = (\lambda - R'')(1 - L) \\
\frac{V''_-}{C'_+} &= (\lambda - (1 - E)R'')L \tag{S13}
\end{aligned}$$

166 Substituting these into the for the appropriate ratios, we obtain:

$$\begin{aligned}
\frac{B(1 - Lp_{\text{in}}) + \left(\frac{N'_+}{T'_+} \frac{N''_-}{N'_+} + \frac{C'_+}{T'_+} \frac{U''_-}{C'_+} \right)}{BLp_{\text{in}} + \frac{C'_+}{T'_+} \frac{V''_-}{C'_+}} &= \frac{B(1 - Lp_{\text{in}}) + \left(\frac{1 - p_{\text{in}}}{1 - LEp_{\text{in}}} (\lambda - R'') + \frac{(1 - LE)p_{\text{in}}}{1 - LEp_{\text{in}}} (\lambda - R'')(1 - L) \right)}{BLp_{\text{in}} + \frac{(1 - LE)p_{\text{in}}}{1 - LEp_{\text{in}}} (\lambda - (1 - E)R'')L} \\
\frac{N'_+ + U'_+ + N_- + U_-}{V'_+ + V_-} &= \frac{B(1 - Lp_{\text{in}}) + \left(1 - Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}} \right) (\lambda - R'')}{BLp_{\text{in}} + Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} \tag{S14}
\end{aligned}$$

167 As in Section S3.1, for the conventional TNCC assumptions, $\lambda \rightarrow 0$ and $R'' \rightarrow 0$. Again, it is not that
168 the secondary process ceases, but just that recruitment via that route is disallowed. Enforcing those con-
169 straints:

$$\begin{aligned}
\lim_{\lambda \rightarrow 0} \frac{B(1 - Lp_{\text{in}}) + \left(1 - Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}}\right) (\lambda - R'')}{BLp_{\text{in}} + Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} &= \frac{B(1 - Lp_{\text{in}}) + \left(1 - Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}}\right) 0}{BLp_{\text{in}} + Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}} 0} \\
&= \frac{1 - Lp_{\text{in}}}{Lp_{\text{in}}} \tag{S15}
\end{aligned}$$

170 As with test-positives odds, we can refactor in terms of the conventional TNCC limit:

$$\begin{aligned}
\frac{B(1 - Lp_{\text{in}}) + \left(1 - Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}}\right) (\lambda - R'')}{BLp_{\text{in}} + Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} &= \frac{1 - Lp_{\text{in}}}{Lp_{\text{in}}} \frac{B + \frac{1 - Lp_{\text{in}} \frac{1 - LE}{1 - LEp_{\text{in}}}}{1 - Lp_{\text{in}}} (\lambda - R'')}{B + \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} \\
&= \frac{1 - Lp_{\text{in}}}{Lp_{\text{in}}} \frac{B + \frac{1 - LEp_{\text{in}} - Lp_{\text{in}}(1 - LE)}{(1 - Lp_{\text{in}})(1 - LEp_{\text{in}})} (\lambda - R'')}{B + \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} \\
&= \frac{1 - Lp_{\text{in}}}{Lp_{\text{in}}} \frac{B + \frac{1 - Lp_{\text{in}} - LEp_{\text{in}} + L^2 Ep_{\text{in}}}{1 - Lp_{\text{in}} - LEp_{\text{in}} + L^2 Ep_{\text{in}}^2} (\lambda - R'')}{B + \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + ER'')} \tag{S16}
\end{aligned}$$

171 In Eq. S16, we now have only terms that describe the intervention (p_{in} , L , and E) and epidemiology
172 (R'' , λ , and B). Note that this term includes more of the model parameters than the test-positive odds
173 (Eq. S10).

174 S3.3 Relaxing Assumption that Prevented Infections Remain Secondary 175 Recruits

176 Earlier, we assumed that the secondary recruitment process was unperturbed by the study intervention. For
177 a secondary process that is, for example, purely geographical because it concerns a pathogen that is highly
178 asymptomatic (*e.g.*, neighbor-household testing for dengue), this assumption is consistent. Where it becomes
179 less clearly acceptable as a simplification, is if there remains some disease- or symptom-based component to
180 secondary recruitment.

181 In the main text, we focus on a vaccine study for EVD, where the primary process was self-reporting
182 with multiple EVD-like symptoms leading to testing. The secondary process is nominally contact-tracing
183 combined with a high-fever. We assumed that high-fever would almost always be present for the contacts
184 that avoided EVD infection because of the vaccine. In reality there would be some attack rate less than
185 100%.

186 If we define the proportion of people meeting a symptom-based component of the secondary process as α
187 and ignore the zero-bias term (Eq. S15) as a coefficient, then Eq. S16 becomes:

$$\text{test-negative odds} \propto \frac{B + \frac{1 - Lp_{\text{in}} - LEp_{\text{in}} + L^2 Ep_{\text{in}}}{1 - Lp_{\text{in}} - LEp_{\text{in}} + L^2 Ep_{\text{in}}^2} (\lambda - R'')}{B + \frac{1 - LE}{1 - LEp_{\text{in}}} (\lambda - R'' + E\alpha R'')} \tag{S17}$$

188 That is, of all the potential secondary recruits that could be added to test-negatives due to prevention of
189 infection, only some exhibit the additional criteria. Note that there is no impact of relaxing this assumption
190 on the test-positive odds.

191 If $\alpha \rightarrow 1$, *i.e.* everyone meets this extra criteria, we get Eq. S16. As $\alpha \rightarrow 0$, the denominator decreases,
192 increasing the test-negative odds overall, and in turn reducing the estimated efficacy. When $\alpha = 0$:

$$\text{test-negative odds} \propto \frac{B + \frac{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}}{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}^2} (\lambda - R'')}{B + \frac{1-LE}{1-LEp_{\text{in}}} (\lambda - R'')} \quad (\text{S18})$$

193 In the case of $p_{\text{in}} = 1$, this reduces to unity. More generally,

$$\begin{aligned} Lp_{\text{in}} \leq L &\implies 1 - Lp_{\text{in}} \geq 1 - L \implies \frac{1-L}{1-Lp_{\text{in}}} \leq 1 \\ &\implies 1 \geq p_{\text{in}} \frac{1-L}{1-Lp_{\text{in}}} \implies LE \geq LEp_{\text{in}} \frac{1-L}{1-Lp_{\text{in}}} \\ &\implies 1 - LE \leq 1 - LEp_{\text{in}} \frac{1-L}{1-Lp_{\text{in}}} \implies 1 - LE \leq \frac{1-Lp_{\text{in}} - LEp_{\text{in}} - L^2Ep_{\text{in}}}{1-Lp_{\text{in}}} \\ &\implies \frac{1-LE}{1-LEp_{\text{in}}} \leq \frac{1-Lp_{\text{in}} - LEp_{\text{in}} - L^2Ep_{\text{in}}}{(1-LEp_{\text{in}})(1-Lp_{\text{in}})} \\ &\implies B + \frac{1-LE}{1-LEp_{\text{in}}} (\lambda - R'') \leq B + \frac{1-Lp_{\text{in}} - LEp_{\text{in}} - L^2Ep_{\text{in}}}{(1-LEp_{\text{in}})(1-Lp_{\text{in}})} (\lambda - R'') \\ &\implies 1 \leq \frac{B + \frac{1-Lp_{\text{in}}-LEp_{\text{in}}-L^2Ep_{\text{in}}}{(1-LEp_{\text{in}})(1-Lp_{\text{in}})} (\lambda - R'')}{B + \frac{1-LE}{1-LEp_{\text{in}}} (\lambda - R'')} \end{aligned} \quad (\text{S19})$$

194 This means that without any contribution from α , test-negative odds biases increasingly towards underesti-
195 mation as participation decreases. As $\alpha \rightarrow 1$, this effect is counteracted, but can in turn lead to overestimation
196 of efficacy.

197 S4 Total Estimator Bias & Limiting Scenarios

198 Combining the test-positives odds and test-negatives odds:

$$\begin{aligned} \hat{E} &= 1 - \frac{V'_+ + V''_+}{N'_+ + N''_+ + U'_+ + U''_+} \frac{N'_- + U'_- + N''_- + U''_-}{V'_- + V''_-} \\ &= 1 - \frac{Lp_{\text{in}}(1-E)}{1-Lp_{\text{in}}} \left[1 + \frac{ER''}{1+R''(1-LE)} \frac{L(1-p_{\text{in}})}{1-Lp_{\text{in}}} \right]^{-1} \frac{1-Lp_{\text{in}}}{Lp_{\text{in}}} \frac{B + \frac{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}}{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}^2} (\lambda - R'')}{B + \frac{1-LE}{1-LEp_{\text{in}}} (\lambda - R'' + ER'')} \\ &= 1 - (1-E) \left[1 + \frac{ER''}{1+R''(1-LE)} \frac{L(1-p_{\text{in}})}{1-Lp_{\text{in}}} \right]^{-1} \frac{B + \frac{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}}{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}^2} (\lambda - R'')}{B + \frac{1-LE}{1-LEp_{\text{in}}} (\lambda - R'' + ER'')} \end{aligned} \quad (\text{S20})$$

199 This suggests a different factorization:

$$\hat{E} = 1 - (1-E) \left[1 + \frac{E \frac{R''}{\lambda} \frac{\lambda}{B+1} (B+1)}{1 + \frac{R''}{\lambda} \frac{\lambda}{B+1} (B+1)(1-LE)} \frac{L(1-p_{\text{in}})}{1-Lp_{\text{in}}} \right]^{-1} \frac{1 + \frac{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}}{1-Lp_{\text{in}}-LEp_{\text{in}}+L^2Ep_{\text{in}}^2} \frac{\lambda}{B+1} \frac{B+1}{B} \left(1 - \frac{R''}{\lambda}\right)}{1 + \frac{1-LE}{1-LEp_{\text{in}}} \frac{\lambda}{B+1} \frac{B+1}{B} \left(1 - \frac{R''}{\lambda} + E \frac{R''}{\lambda}\right)} \quad (\text{S21})$$

200 We can define secondary test-positive fraction, $p_t = \frac{R''}{\lambda}$, the negative proportion of primary alerts, $f_- = \frac{B}{B+1}$,
201 and relative rate of secondary recruitment to primary recruitment, $\rho = \frac{\lambda}{B+1}$. Noting that $B+1 = (1-f_-)^{-1}$

202 This yields:

$$\hat{E} = 1 - (1 - E) \left[1 + \frac{E \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \frac{L(1 - p_{in})}{1 - Lp_{in}} \right]^{-1} \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1 - p_t(1 - E))} \quad (\text{S22})$$

203 We use this framing for all the main text results. This formulation highlights the important relative values
 204 within the model, while still maintaining terms that can be reasoned about and potentially measured. In
 205 this framing \hat{E} is still a function of six variables, *i.e.* $\{E, L, p_{in}, R'', \lambda, B\}$ versus $\{E, L, p_{in}, p_t, \rho, f_-\}$.

206 In the following subsections, we show limiting conditions for the estimator with respect to the assorted
 207 parameters.

208 S4.1 True efficacy, E , limits

209 As the intervention tends toward either doing nothing ($E \rightarrow 0$) or perfect protection ($E \rightarrow 1$), the estimator
 210 bias tends to vanish.

$$\begin{aligned} \lim_{E \rightarrow 0} \hat{E} &= 1 - [1]^{-1} \frac{1 + \frac{1-Lp_{in}}{1-Lp_{in}} \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{\rho}{f_-} (1 - p_t)} = 1 - 1 = 0 \\ \lim_{E \rightarrow 1} \hat{E} &= 1 - 0(\dots) = 1 \end{aligned} \quad (\text{S23})$$

211 S4.2 Participating fraction, p_{in} , limits

$$\begin{aligned} \lim_{p_{in} \rightarrow 0} \hat{E} &= 1 - (1 - E) \left[1 + \frac{LE \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \right]^{-1} \frac{1 + \frac{1}{1} \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{1-LE}{1} \frac{\rho}{f_-} (1 - p_t(1 - E))} \\ &= 1 - (1 - E) \left[\frac{1 + \frac{p_t \rho}{1-f_-} (1 - LE)}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} + \frac{LE \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \right]^{-1} \frac{1 + \frac{\rho}{f_-} (1 - p_t)}{1 + (1 - LE) \frac{\rho}{f_-} (1 - p_t(1 - E))} \\ &= 1 - (1 - E) \left[\frac{1 + \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \right]^{-1} \frac{1 + \frac{\rho}{f_-} (1 - p_t)}{1 + (1 - LE) \frac{\rho}{f_-} (1 - p_t(1 - E))} \\ &= 1 - (1 - E) \left[\frac{1 + \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \right]^{-1} \frac{1 + \frac{\rho}{f_-} (1 - p_t)}{1 + (1 - LE) \frac{\rho}{f_-} (1 - p_t(1 - E))} \\ &= 1 - (1 - E) \frac{1 + \frac{p_t \rho}{1-f_-} (1 - LE)}{1 + \frac{p_t \rho}{1-f_-}} \frac{1 + \frac{\rho}{f_-} (1 - p_t)}{1 + (1 - LE) \frac{\rho}{f_-} (1 - p_t(1 - E))} \end{aligned} \quad (\text{S24})$$

$$\begin{aligned}
\lim_{p_{\text{in}} \rightarrow 1} \hat{E} &= 1 - (1 - E) [1]^{-1} \frac{B + [1 + \frac{L}{1-L} 0](\lambda - R'')}{B + (\lambda - R'' + ER'')} \\
&= 1 - (1 - E) \left[1 + 0 \frac{E \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \right]^{-1} \frac{1 + \frac{1-L-LE+L^2E}{1-L-LE+L^2E} \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{1-L-E}{1-LE} \frac{\rho}{f_-} (1 - p_t (1 - E))} \\
&= 1 - (1 - E) \frac{1 + \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{\rho}{f_-} (1 - p_t (1 - E))} \\
&= 1 - (1 - E) \frac{B + \lambda - R''}{B + \lambda - R'' + ER''} \\
&= 1 - (1 - E) \frac{1 - \frac{R''}{B+\lambda}}{1 - \frac{R''}{B+\lambda} (1 - E)} \tag{S25}
\end{aligned}$$

212 The final factorization in Eq. S25 shows that, in the limit of perfect alignment of participation and recruit-
213 ment, the intervention coverage, L , is removed, and the bias depends only on the true efficacy, E , and a
214 combination of epidemiological parameters: $\frac{R''}{B+\lambda}$. This relationship can be inverted; which allows us to
215 determine the true efficacy in terms of the estimator value and other parameters:

$$\begin{aligned}
1 - \hat{E} &= (1 - E) \frac{1 - \frac{R''}{B+\lambda}}{1 - \frac{R''}{B+\lambda} (1 - E)} \\
(1 - \hat{E}) \left(1 - \frac{R''}{B+\lambda} (1 - E) \right) &= (1 - E) \left(1 - \frac{R''}{B+\lambda} \right) \\
(1 - \hat{E}) &= (1 - E) \left[1 - \frac{R''}{B+\lambda} + (1 - \hat{E}) \frac{R''}{B+\lambda} \right] \\
(1 - E) &= (1 - \hat{E}) \left[1 - \hat{E} \frac{R''}{B+\lambda} \right]^{-1} \\
E &= 1 - (1 - \hat{E}) \left[1 - \hat{E} \frac{R''}{B+\lambda} \right]^{-1} \\
E - \hat{E} &= (1 - \hat{E}) \left(1 - \left[1 - \hat{E} \frac{R''}{B+\lambda} \right]^{-1} \right) \\
E - \hat{E} &= -(1 - \hat{E}) \left(\frac{\hat{E} \frac{R''}{B+\lambda}}{1 - \hat{E} \frac{R''}{B+\lambda}} \right) \tag{S26}
\end{aligned}$$

216 $\frac{R''}{B+\lambda}$ corresponds to a potentially measurable quantity; recall that in the model, R'' is the expected number of
217 additional test-positives that are identified via the secondary process in a group without the intervention, and
218 $B + \lambda$ is all the other tests (primary negatives and all secondary tests) per primary test-positive. Thus, $\frac{R''}{B+\lambda}$
219 is the fraction of secondary test-positives out of all non-index case-finding tests, when measuring in a non-
220 intervention group. This value could be estimated from a comparable population without the intervention,
221 or an upper limit could be estimated from data within the study population itself: the expected number
222 of secondary cases in the intervention population is $R''(1 - LE)$, so the measured non-primary test-positive
223 fraction would be reduced maximally by a factor $(1 - L)$ when the efficacy is perfect. Note that because
224 this factor only has $B + \lambda$, we do not need to distinguish primary versus secondary test-negatives. If a
225 test-negative from λ was mistakenly assigned to B (or *vice versa*), that would not change this factor.

226 Thus, if a study were able to achieve $p_{\text{in}} \approx 1$, it only need to be able to distinguish between primary and
227 secondary test-positives to correctly bound the estimator.

228 **S4.3 Secondary Case Recruitment, p_t , limits**

229 If we consider secondary *case* limits, that is what proportion of secondary recruiting is test-positive, then we
 230 obtain:

$$\begin{aligned}
 \lim_{p_t \rightarrow 0} \hat{E} &= 1 - (1 - E) \left[1 + \frac{0}{1+0} \frac{L(1-p_{in})}{1-Lp_{in}} \right]^{-1} \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-}}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-}} \\
 &= 1 - (1 - E) \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-}}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-}} \\
 \lim_{p_t \rightarrow 1} \hat{E} &= 1 - (1 - E) \left[1 + \frac{E \frac{\rho}{1-f_-}}{1 + \frac{\rho}{1-f_-} (1-LE)} \frac{L(1-p_{in})}{1-Lp_{in}} \right]^{-1} \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} 0}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} E} \\
 &= 1 - (1 - E) \left[1 + \frac{E \frac{\rho}{1-f_-}}{1 + \frac{\rho}{1-f_-} (1-LE)} \frac{L(1-p_{in})}{1-Lp_{in}} \right]^{-1} \frac{1}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} E} \tag{S27}
 \end{aligned}$$

231 One way to interpret $p_t \rightarrow 1$ is that transmission probability (conditional on high risk contact) is going up.
 232 Another way to think about it is λ coming down to meet R'' ; *i.e.*, the decision about whether to test a
 233 contact or not becoming more accurately linked to whether they were infected.

234 **S4.4 Lower Limit on Secondary Relative Recruiting, ρ**

235 As primary recruiting increasingly outweighs secondary recruiting (including both increasing primary re-
 236 cruitment and disallowing secondary recruitment), $\rho \rightarrow 0$. In this limit:

$$\begin{aligned}
 \lim_{\rho \rightarrow 0} \hat{E} &= 1 - (1 - E) \left[1 + \frac{0}{1+0} \frac{L(1-p_{in})}{1-Lp_{in}} \right]^{-1} \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{0}{f_-} (1-p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{0}{f_-} (1-p_t(1-E))} \\
 &= 1 - (1 - E) [1]^{-1} \frac{1}{1} = E \tag{S28}
 \end{aligned}$$

237 Thus, for sufficiently high rate of primary recruitment leading to test-negatives, the bias goes to 0.

238 **S5 Translation of Limits to Recruitment Constraints**

239 **S5.1 Attempting to Limit Recruitment to Participants Only**

240 Figures S2-S10 show the general response of the estimator to varying factors in the model. Each plot shows
 241 $p_{in} \in \{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$ (columns) and $p_t \in \{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$ (rows). Each plot
 242 shows one of the combinations of $\rho \in \{1/9, 1/3, 1\}$ and $f_- \in \{0.5, 0.75, 0.9\}$; ρ is indicated at the top of each
 243 plot, f_- on the right side.

244 These plots show some trends under specific conditions. In general, increasing participation decreases bias
 245 range, though not absolutely (*e.g.*, $\rho = 1/9, p_t = 1/4, f_- = 0.75$). Increasing coverage can shift bias towards
 246 underestimation or overestimation, depending p_t .

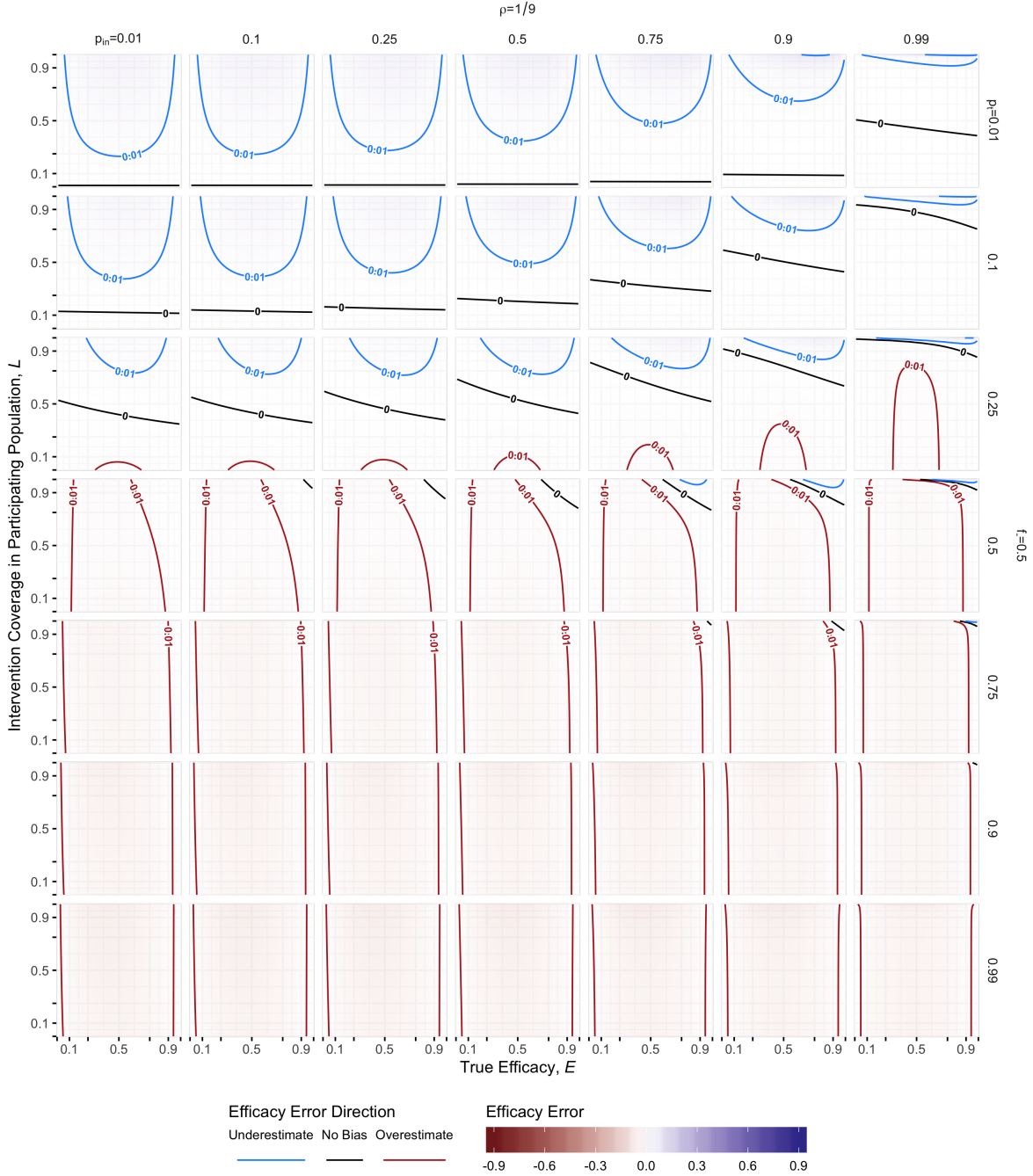


Figure S2: **General Bias Sensitivity, 1 of 9:** These series of plots show general sensitivity of the TNCC estimator to all of the model parameters. For each plot, the rightmost column corresponds to very high (99%) participation fraction, which indicates the minimal bias surface when the study manages to maximize participation. Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.5$.*

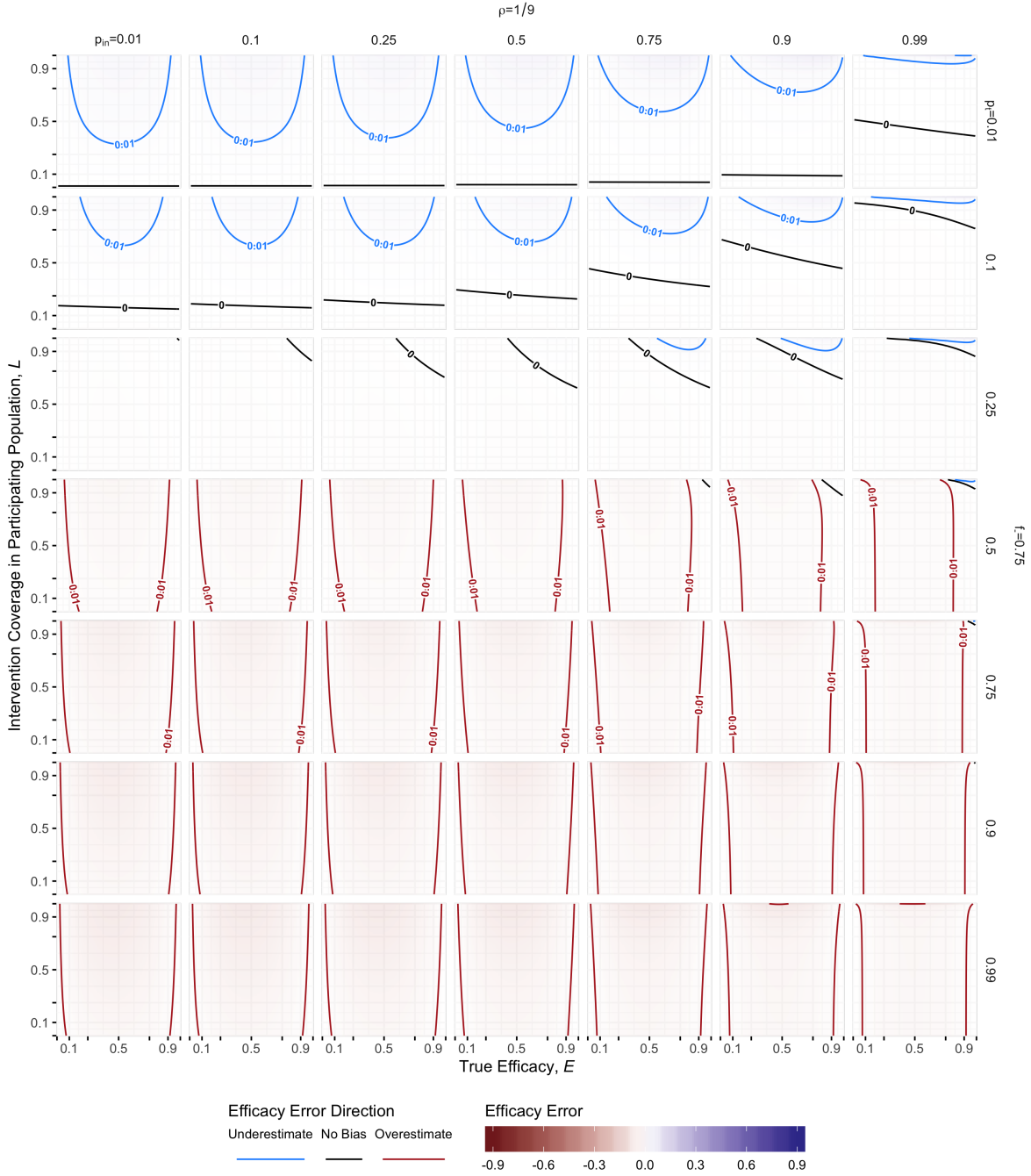


Figure S3: **General Bias Sensitivity. 2 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.75$.*

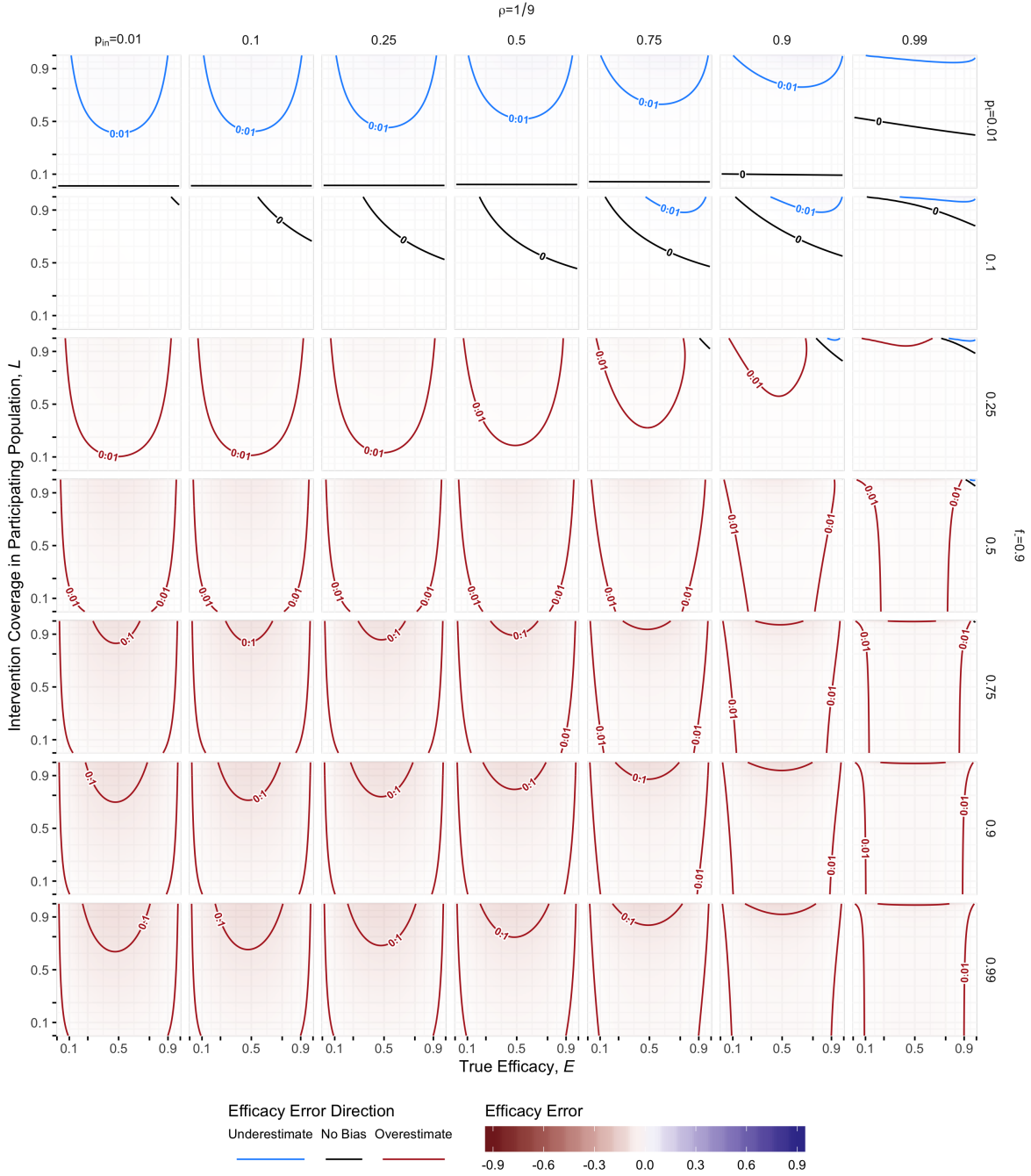


Figure S4: **General Bias Sensitivity. 3 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.9$.*

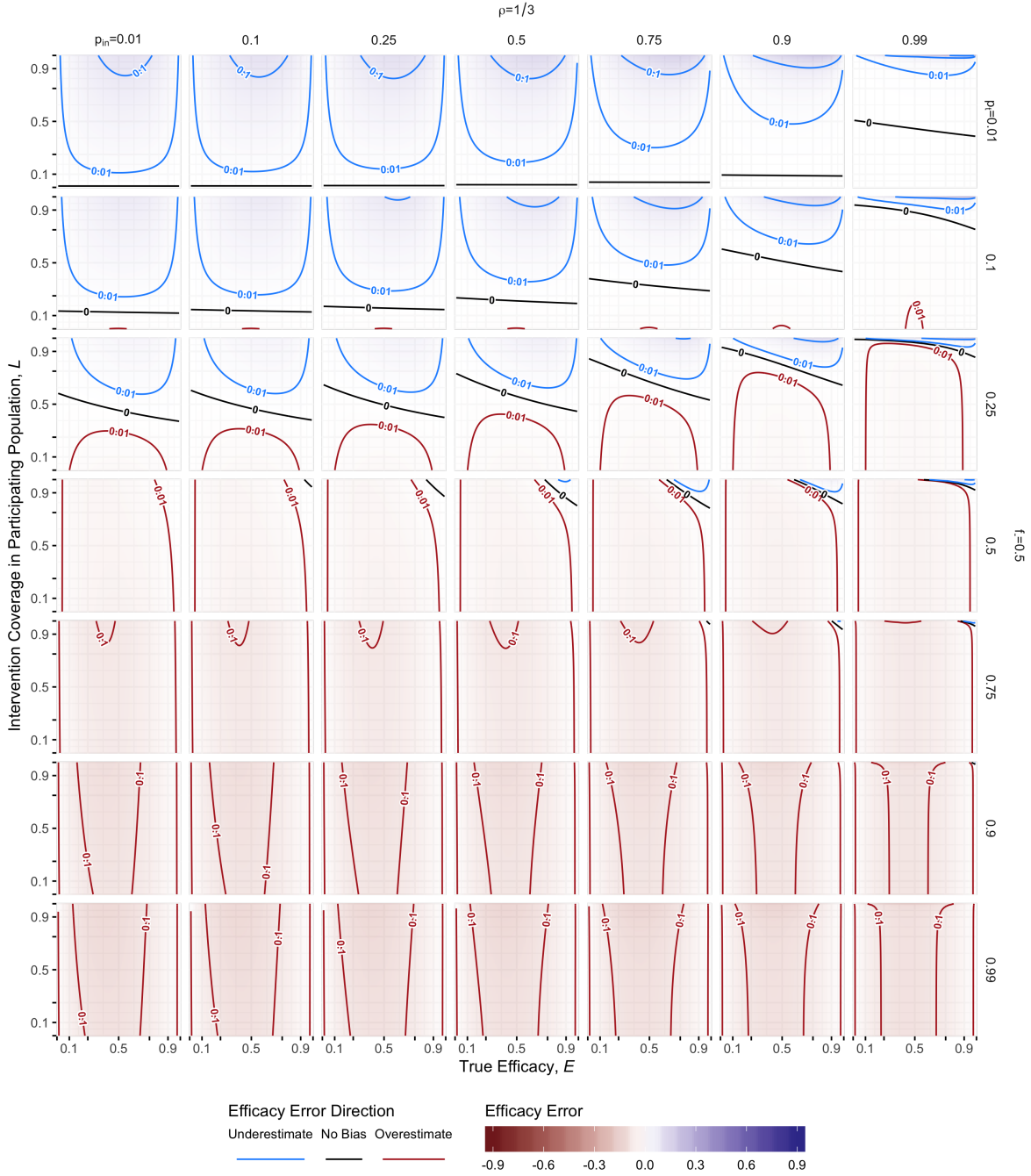


Figure S5: **General Bias Sensitivity. 4 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.5$.*

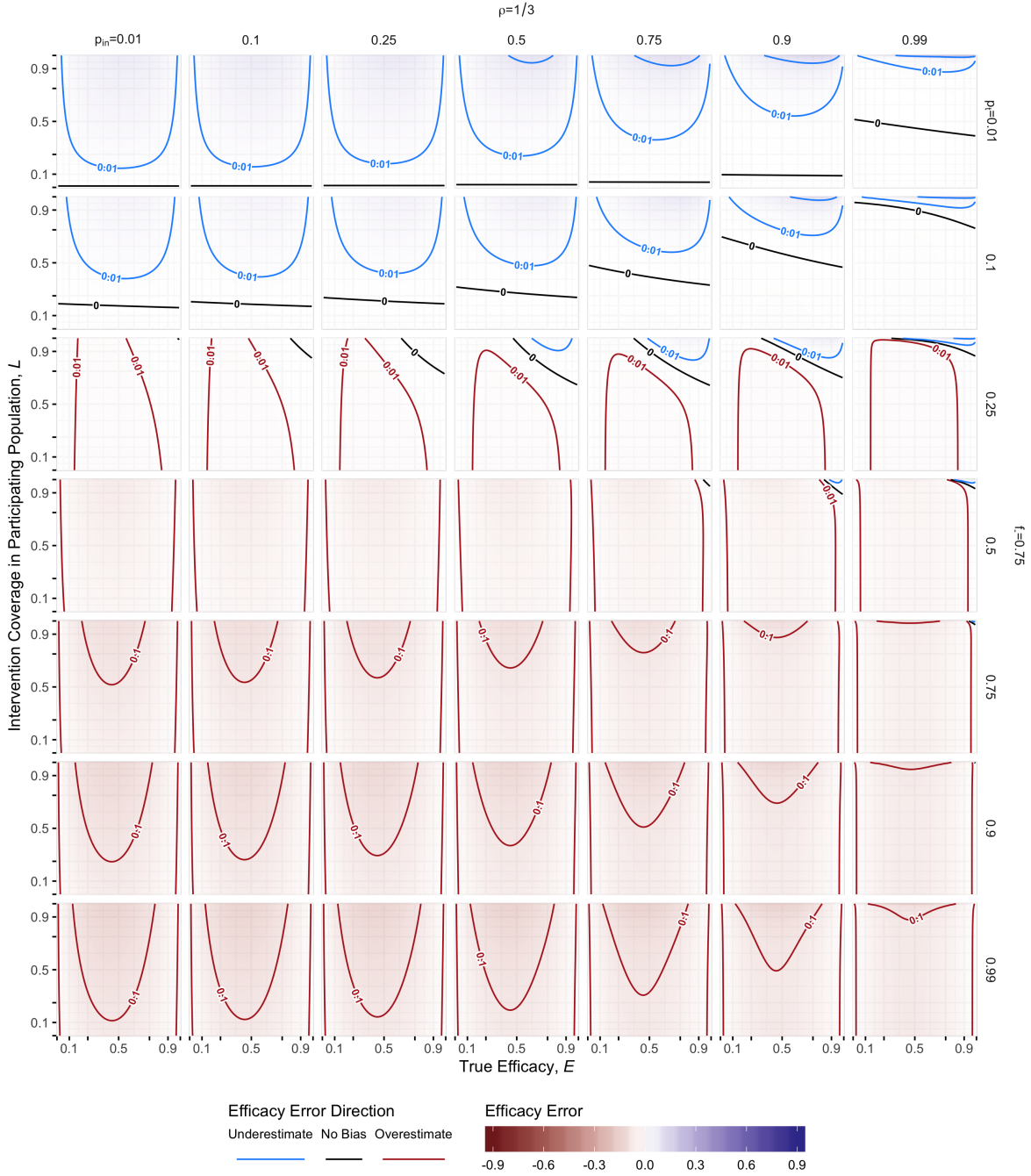


Figure S6: **General Bias Sensitivity. 5 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.75$.*

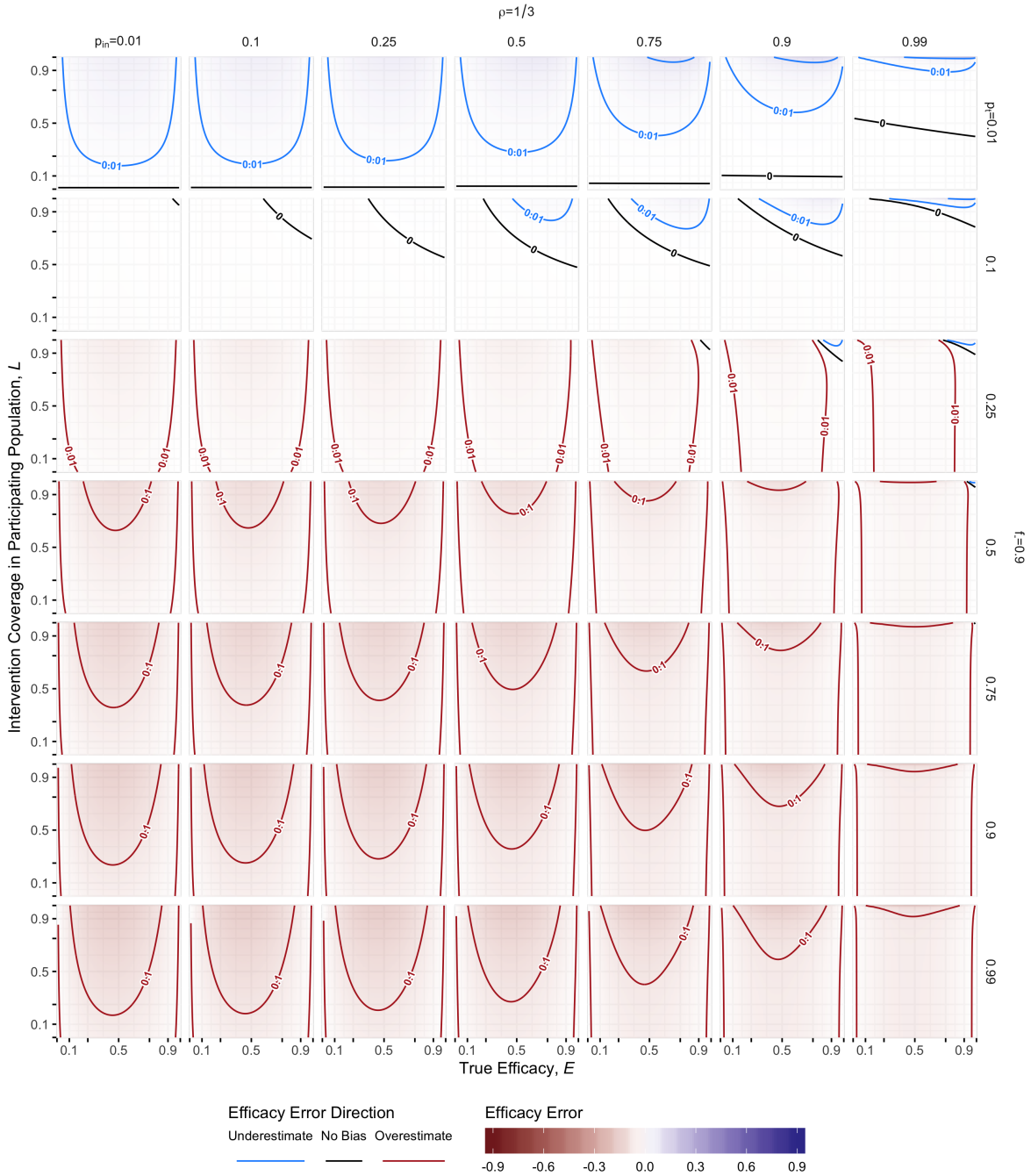


Figure S7: **General Bias Sensitivity. 6 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.9$.*

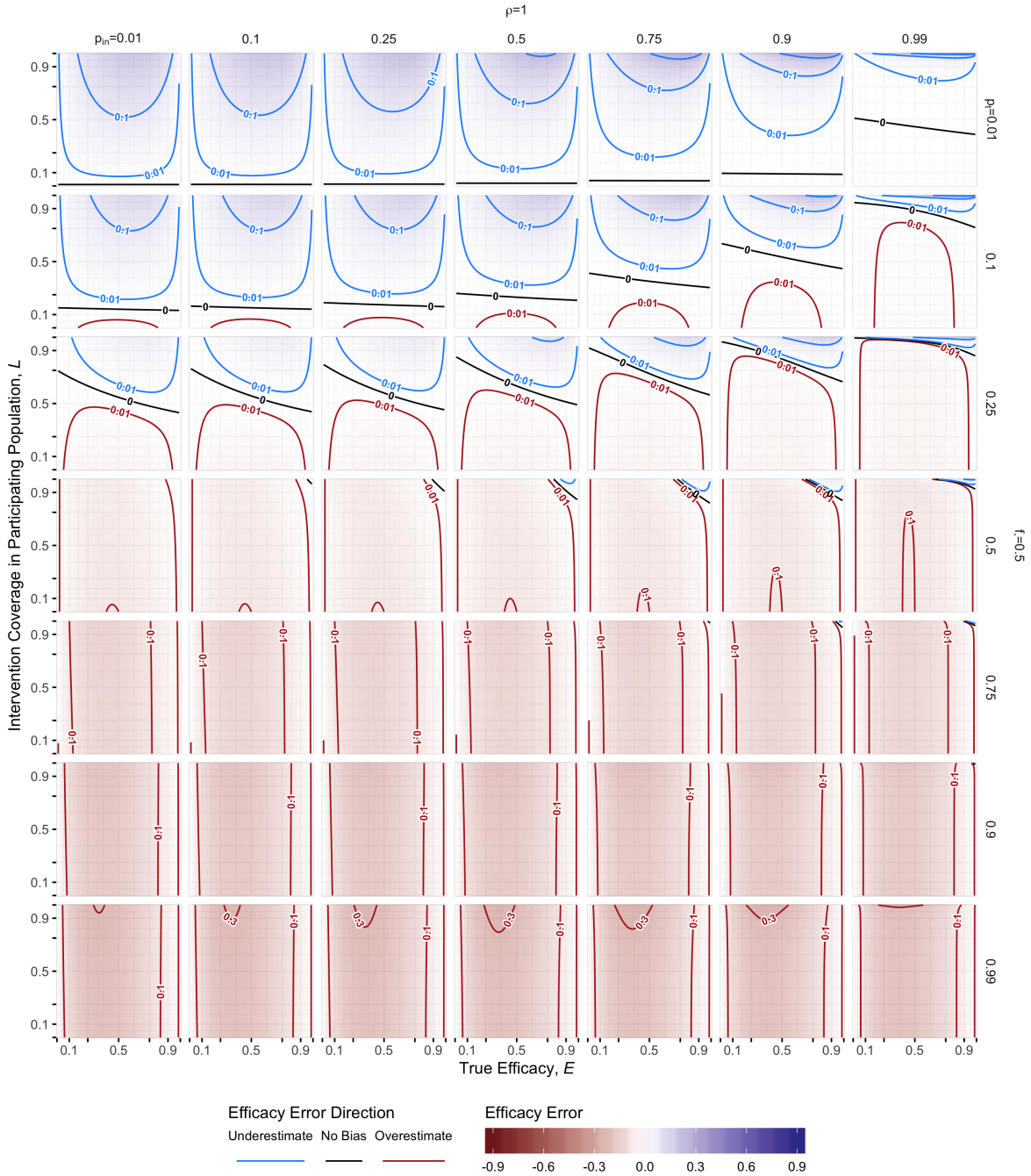


Figure S8: **General Bias Sensitivity. 7 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.5$.*

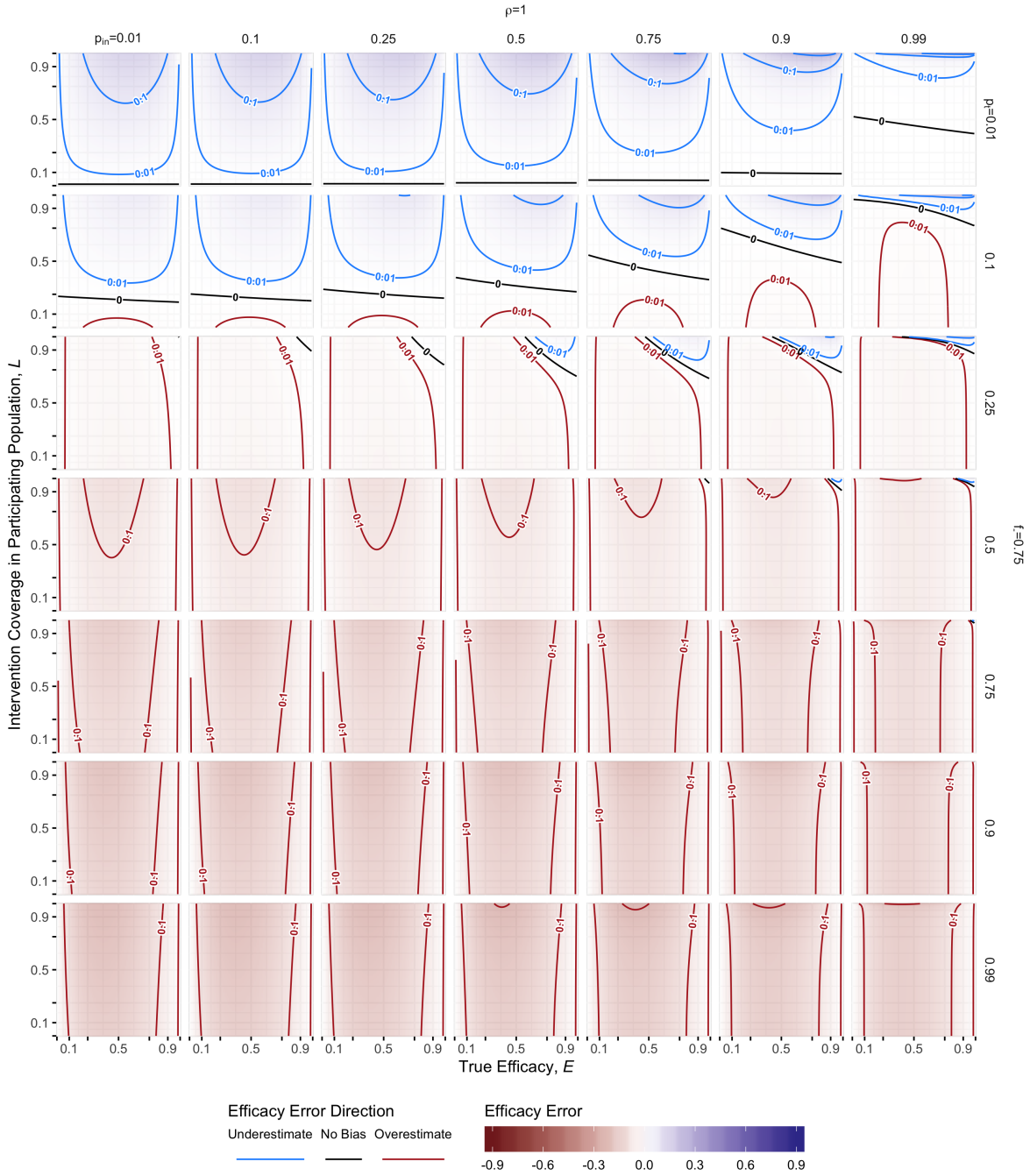


Figure S9: **General Bias Sensitivity. 8 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.75$.*

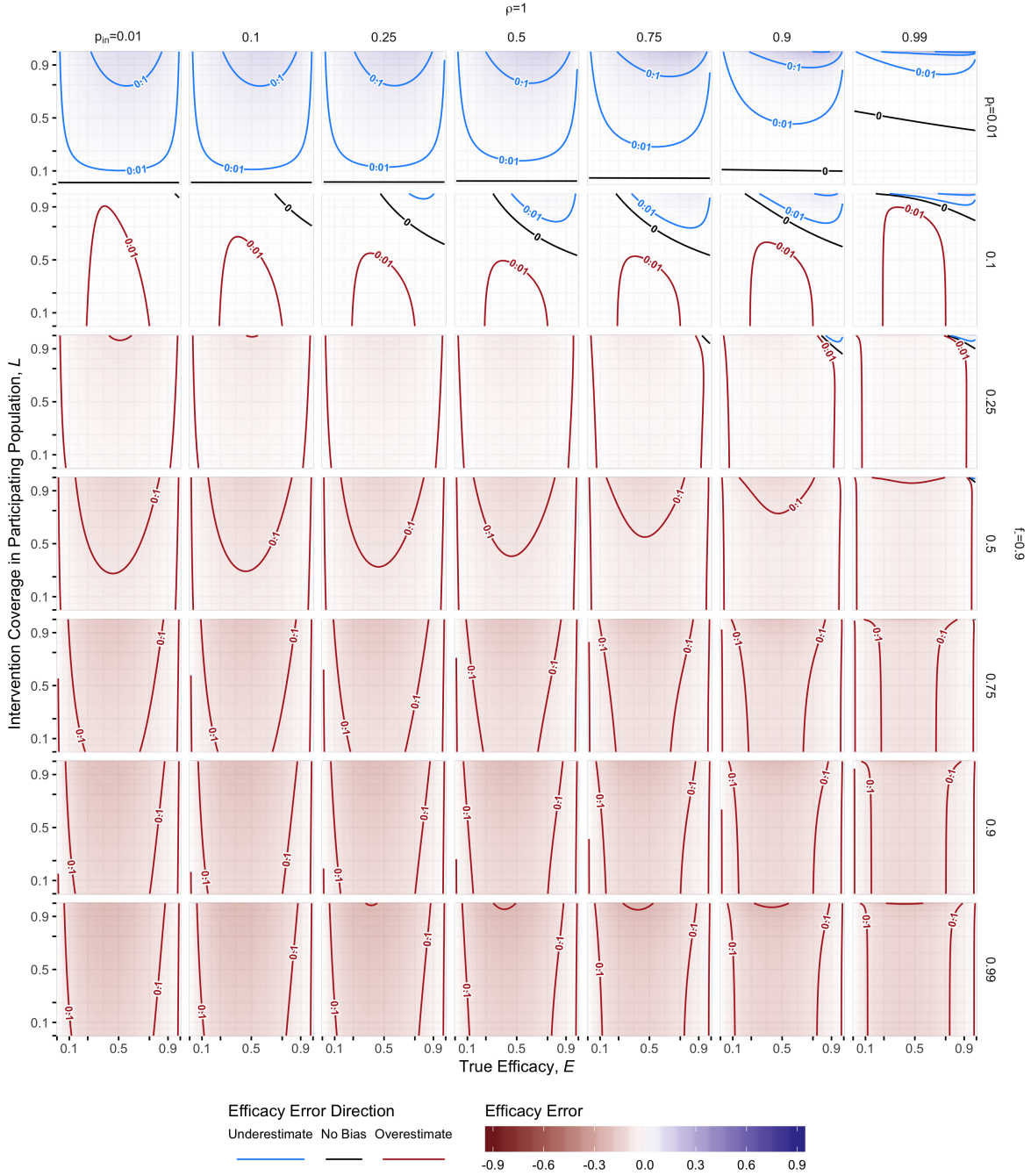


Figure S10: **General Bias Sensitivity.**: Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.9$.*

247 **S5.2 Attempting to Limit to Primary Recruitment Only**

248 An alternative approach to controlling the bias is to restrict to primary recruitment only. If we assume
 249 that the study excludes secondary recruitment perfectly for test-positives (*e.g.* because they are extensively
 250 monitored) but incompletely excludes secondary recruitment for test-negatives (*e.g.* because data for them
 251 is incomplete) then the full estimator equation:

$$\hat{E} = 1 - (1 - E) \left[1 + \frac{E \frac{p_t \rho}{1-f_-}}{1 + \frac{p_t \rho}{1-f_-} (1 - LE)} \frac{L(1 - p_{in})}{1 - Lp_{in}} \right]^{-1} \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} (1 - p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1 - p_t(1 - E))}$$

252 will lose the test-positive bias contribution, because it goes to 1 (note that for any non-zero p_t , this term is
 253 less than 1):

$$\left[1 + \frac{E \frac{0\rho}{1-f_-}}{1 + \frac{0\rho}{1-f_-} (1 - LE)} \frac{L(1 - p_{in})}{1 - Lp_{in}} \right]^{-1} = 1 \quad (\text{S29})$$

254 So the overall bias becomes:

$$\hat{E} = 1 - (1 - E) \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\beta\rho}{f_-} (1 - p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\beta\rho}{f_-} (1 - p_t(1 - E))} \quad (\text{S30})$$

255 where $\beta \in (0, 1)$ is the exclusion failure probability; $\beta = 0$ is perfect exclusion of secondary test-negatives (in
 256 which case the test-negative term also reduces to 1) while $\beta = 1$ implies that no secondary test-negatives are
 257 excluded. Note that this factor is simply reducing ρ in the test-negative odds term. Thus, we can drop β
 258 and instead reduce the range we consider for ρ . The error expression for this scenario is:

$$\begin{aligned}
E - \hat{E} &= E - 1 + (1 - E) \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} (1-p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= -(1-E) \left[1 - \frac{1 + \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} (1-p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1-p_t(1-E))} \right] \\
&= -(1-E) \left[\frac{\frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1-p_t(1-E)) - \frac{1-Lp_{in}-LEp_{in}+L^2Ep_{in}}{1-Lp_{in}-LEp_{in}+L^2Ep_{in}^2} \frac{\rho}{f_-} (1-p_t)}{1 + \frac{1-LE}{1-LEp_{in}} \frac{\rho}{f_-} (1-p_t(1-E))} \right] \\
&= \frac{-(1-E) \frac{\rho}{f_-} (1-Lp_{in})(1-LE)(1-p_t(1-E)) - (1-Lp_{in}-LEp_{in}+L^2Ep_{in})(1-p_t)}{(1-Lp_{in})(1-LEp_{in}) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= \frac{-(1-E) \frac{\rho}{f_-} (1-Lp_{in}-LE+L^2Ep_{in})(1-p_t(1-E)) - (1-Lp_{in}-LEp_{in}+L^2Ep_{in})(1-p_t)}{(1-Lp_{in})(1-LEp_{in}) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= \frac{-(1-E) \frac{\rho}{f_-} (1-Lp_{in}-LE+L^2Ep_{in})Ep_t + -LE(1-p_t) + LEp_{in}(1-p_t)}{(1-Lp_{in})(1-LEp_{in}) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= \frac{-(1-E) \frac{\rho}{f_-} (1-Lp_{in}-LE+L^2Ep_{in})Ep_t - LE(1-p_t)(1-p_{in})}{(1-Lp_{in})(1-LEp_{in}) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= \frac{-(1-E)E \frac{\rho}{f_-} (1-Lp_{in})(1-LE)p_t - L(1-p_t)(1-p_{in})}{(1-Lp_{in})(1-LEp_{in}) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))}
\end{aligned} \tag{S31}$$

259 Because p_{in} remains in the equation, in multiple places, the participating fraction plays an important role
260 in bias. If participation is high, and for the case of $p_{in} \rightarrow 1$, then we can then recover:

$$\begin{aligned}
\lim_{p_{in} \rightarrow 1} E - \hat{E} &= \frac{-(1-E)E \frac{\rho}{f_-} (1-L)(1-LE)p_t - L(1-p_t)(1-1)}{(1-L)(1-LE) + (1-LE) \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= -(1-E)E \frac{\rho}{f_-} \frac{p_t}{1 + \frac{\rho}{f_-} (1-p_t(1-E))} \\
&= -(1-E)E \frac{p_t}{1 + \frac{f_-}{\rho} - p_t(1-E)}
\end{aligned} \tag{S32}$$

261 which is the same error as if we had just been able to control $p_{in} \rightarrow 1$. This suggests that if the study
262 could mostly manage heterogeneity in the recruited population, but likely could not exclude secondary
263 test-negatives, then including all secondary cases does not affect the magnitude of bias.

264 Figures S11-S19 show the response of the estimator to varying factors when secondary test-
265 positives are excluded. Each plot shows $p_{in} \in \{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$ (columns) and $p_t \in$
266 $\{0.01, 0.1, 0.25, 0.5, 0.75, 0.9, 1\}$ (rows). Each plot shows one of the combinations of $\rho \in \{1/9, 1/3, 1\}$ and
267 $f_- \in \{0.5, 0.75, 0.9\}$; ρ is indicated at the top of each plot, f_- on the right side.

268 In general, these plots show the same trends as Figures S2-S10, with lower bias magnitude and tendency to
269 shift towards underestimation.

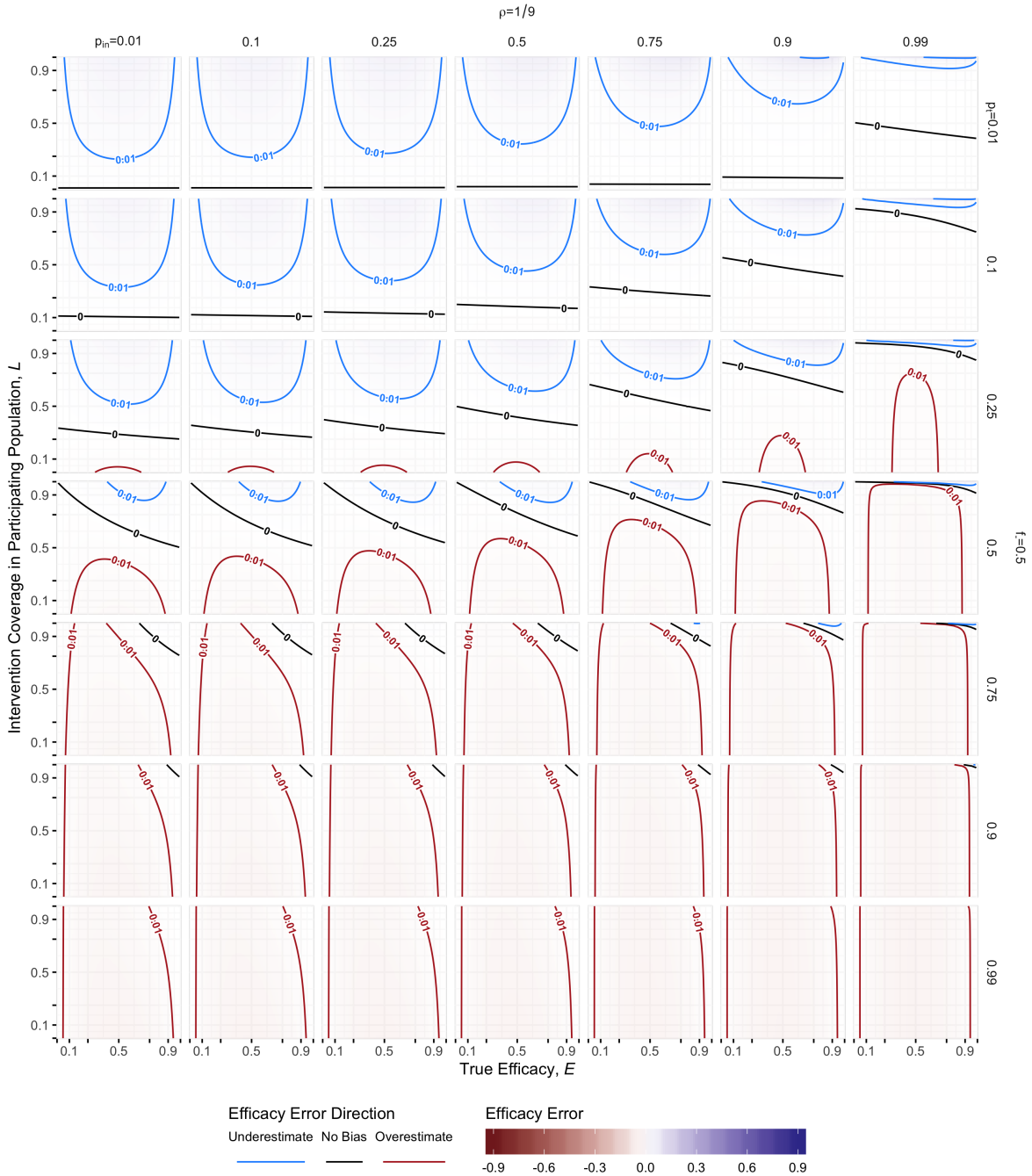


Figure S11: **Bias Sensitivity Without Secondary Test-Positives, 1 of 9:** These series of plots show sensitivity of the TNCC estimator which excludes secondary test-positives to all of the model parameters. For each plot, the rightmost column corresponds to very high (99%) participation fraction, which indicates the minimal bias surface when the study manages to maximize participation. Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.5$.*

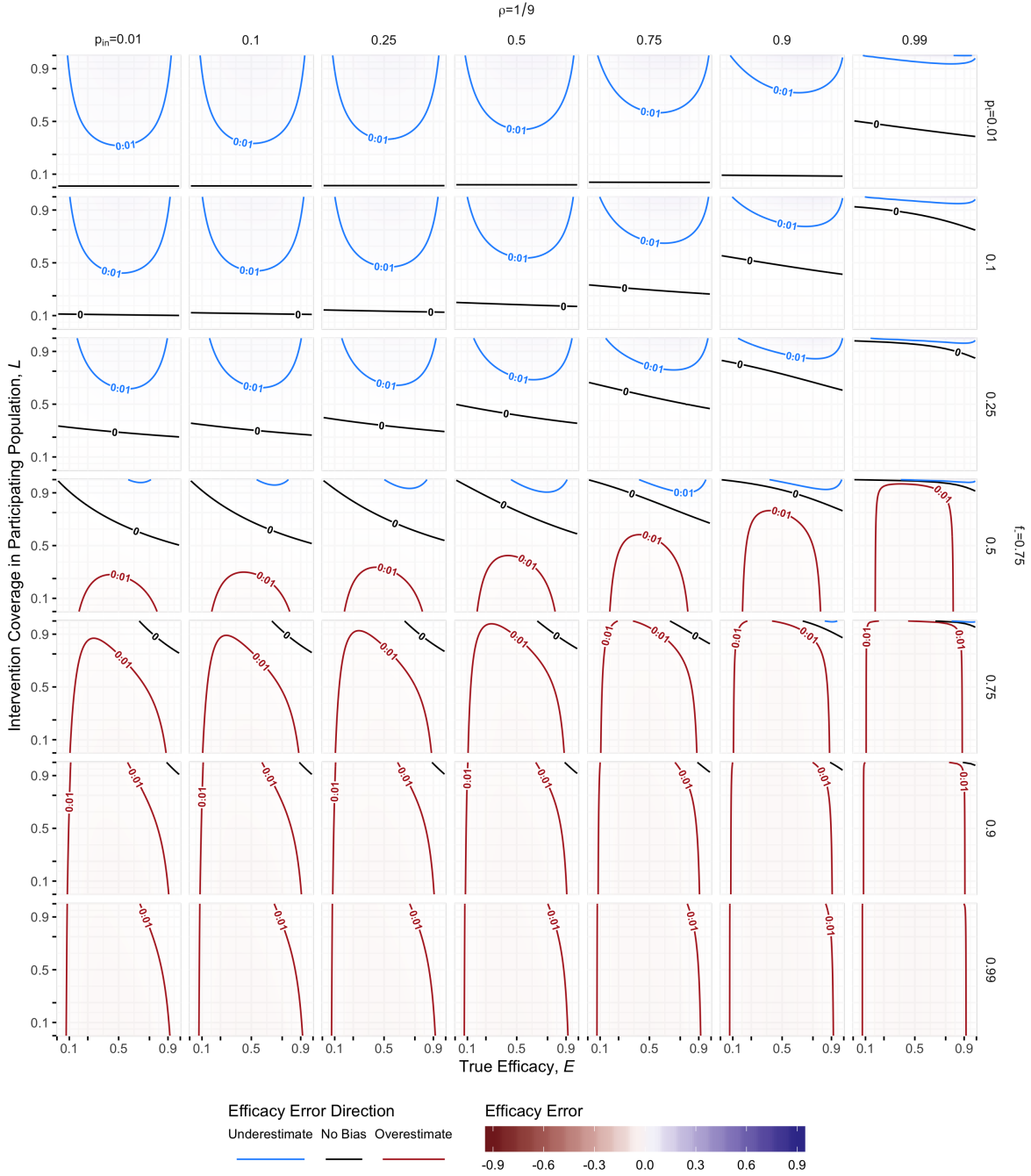


Figure S12: **Bias Sensitivity Without Secondary Test-Positives. 2 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.75$.*

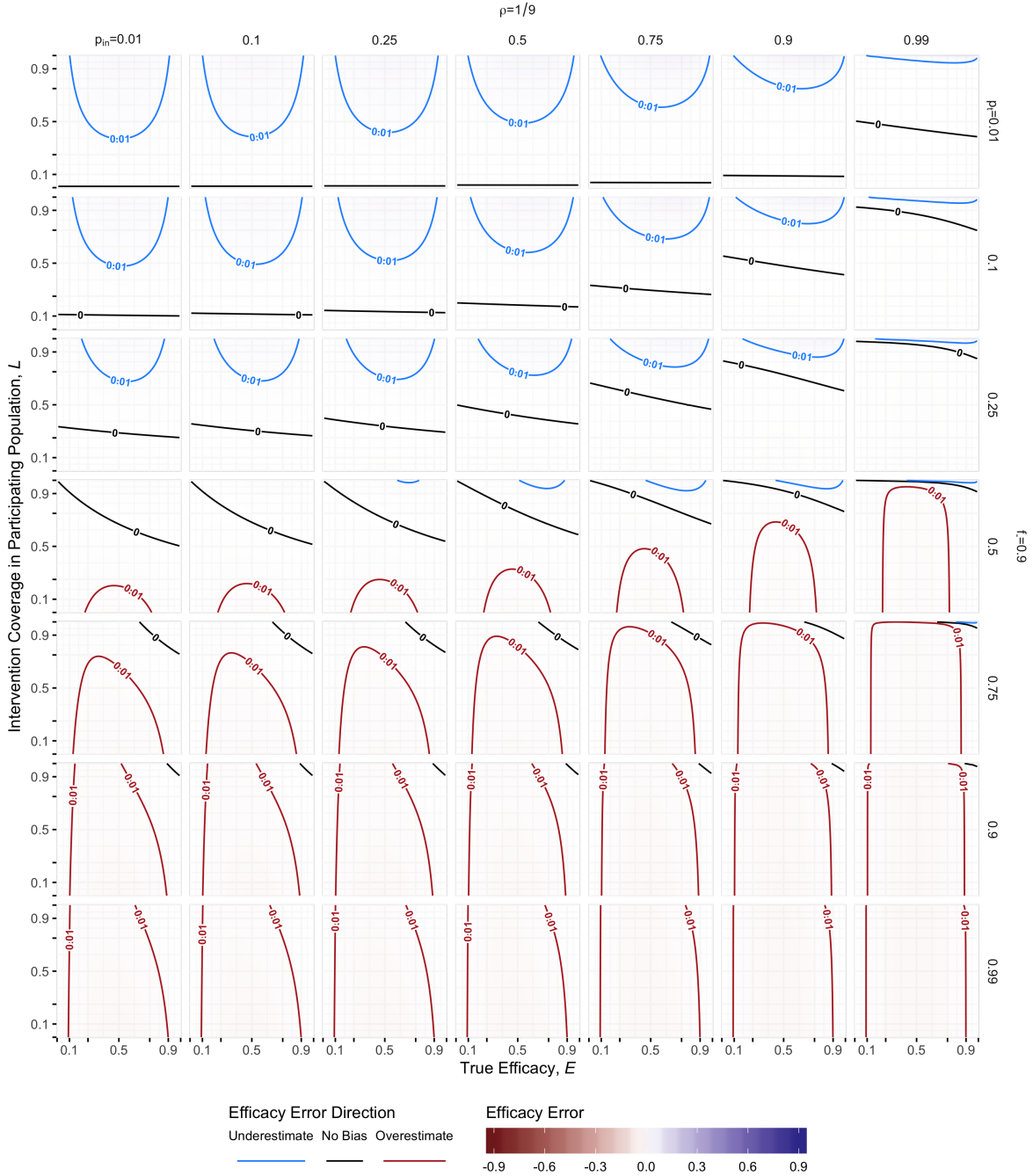


Figure S13: **Bias Sensitivity Without Secondary Test-Positives. 3 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/9$ and $f_- = 0.9$.*

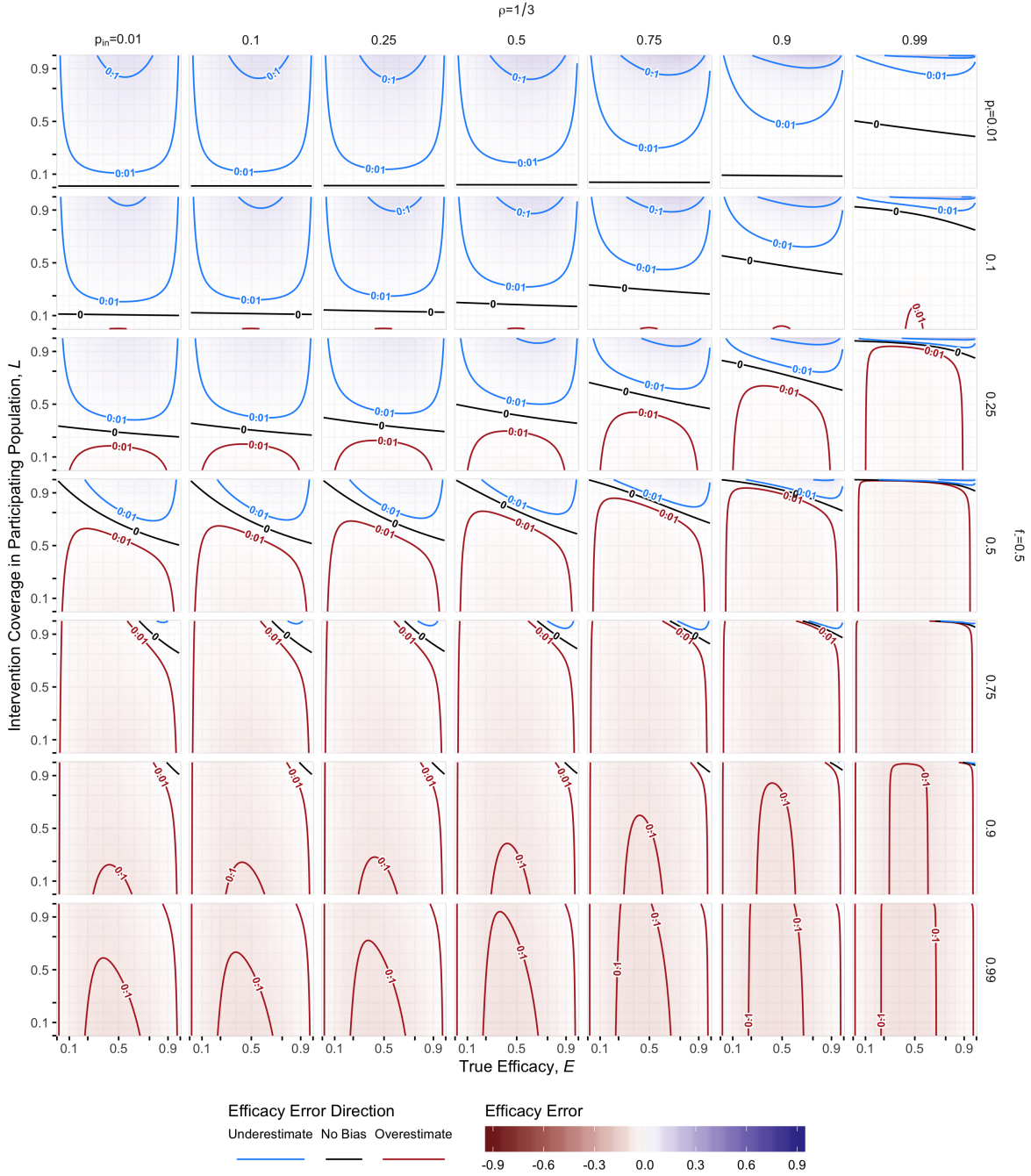


Figure S14: **Bias Sensitivity Without Secondary Test-Positives.** 4 of 9: Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.5$.*

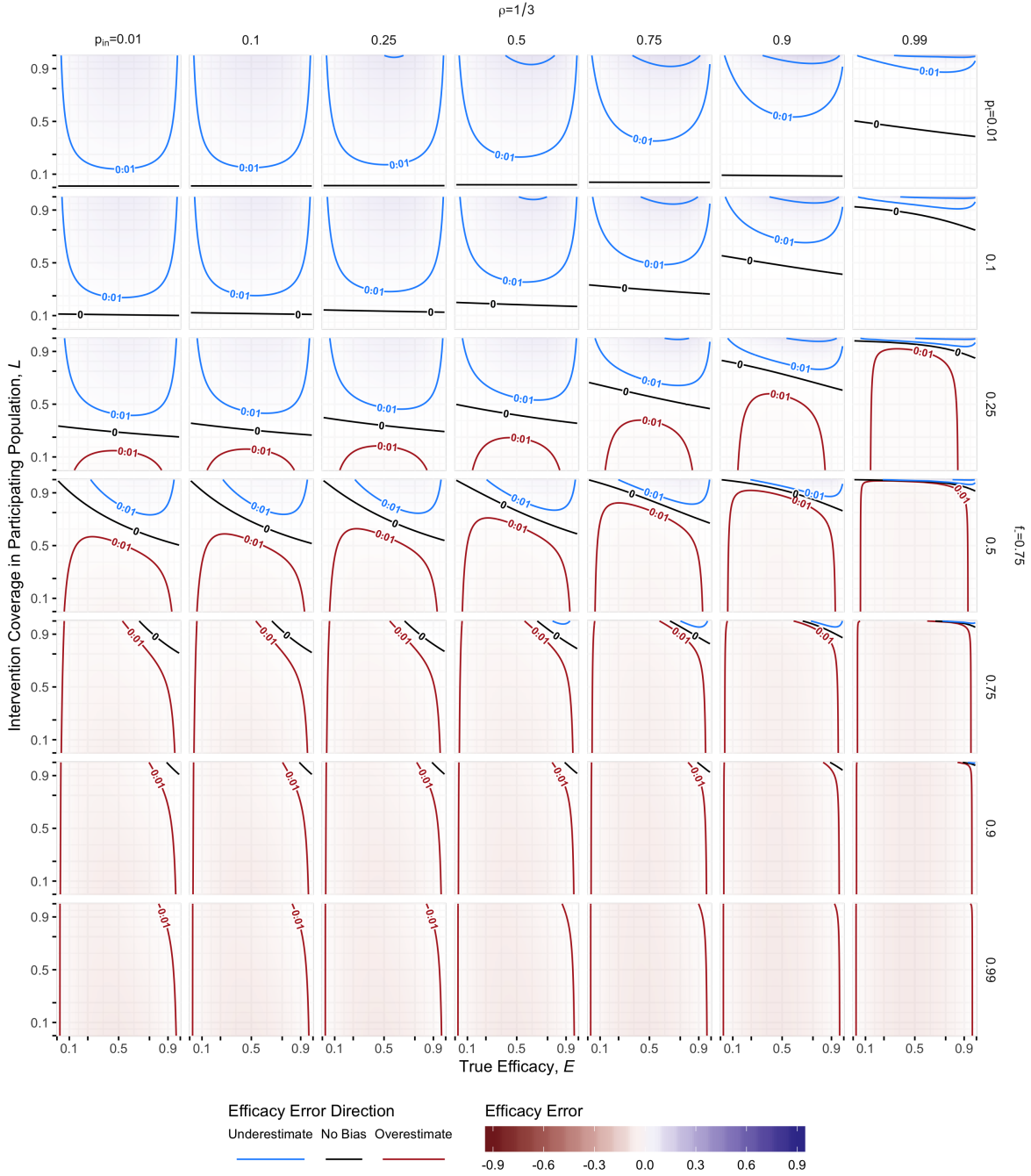


Figure S15: **Bias Sensitivity Without Secondary Test-Positives.** 5 of 9: Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.75$.*

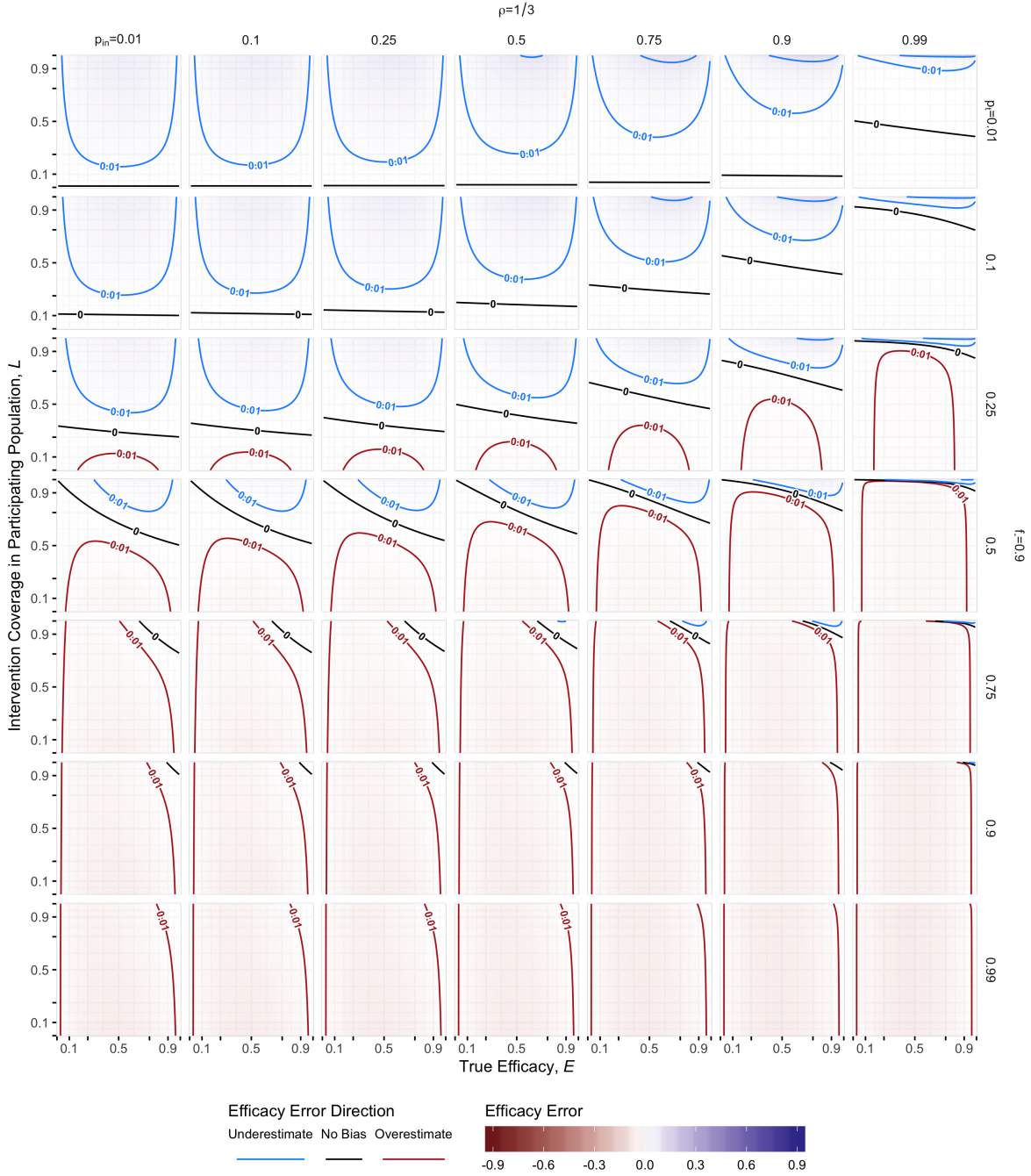


Figure S16: **Bias Sensitivity Without Secondary Test-Positives. 6 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1/3$ and $f_- = 0.9$.*

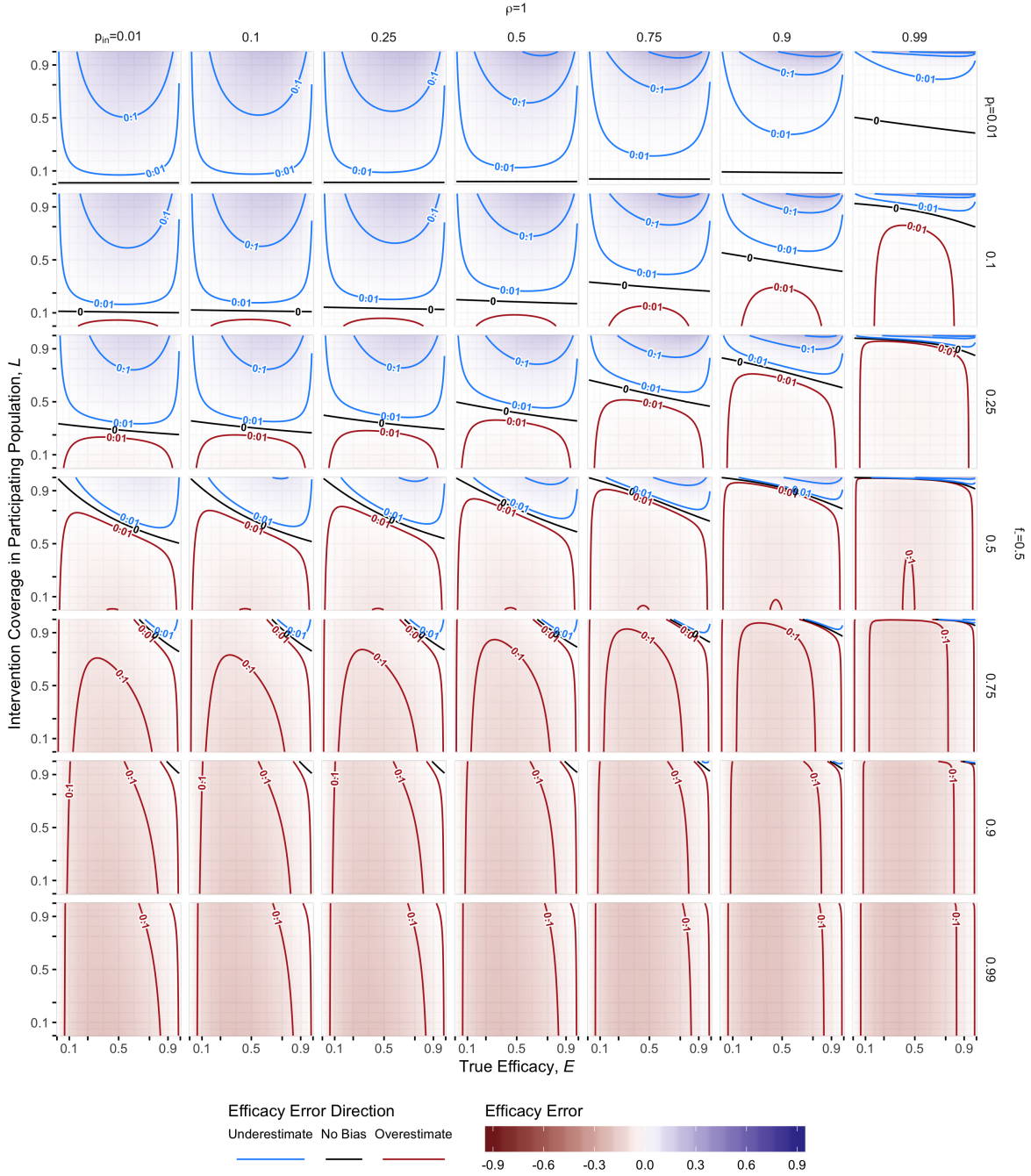


Figure S17: **Bias Sensitivity Without Secondary Test-Positives.** **7 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expect fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.5$.*

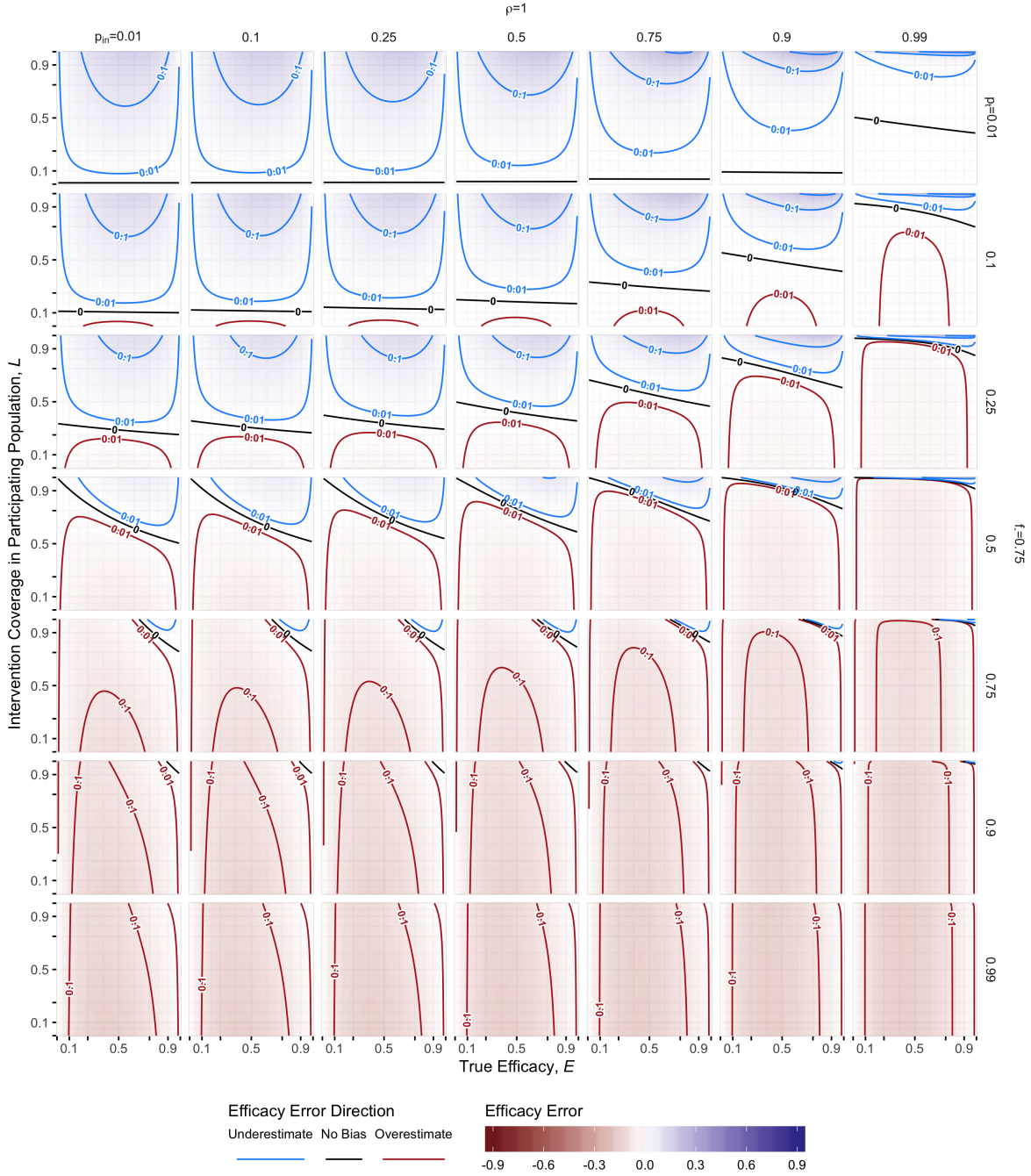


Figure S18: **Bias Sensitivity Without Secondary Test-Positives.** **8 of 9:** Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.75$.*

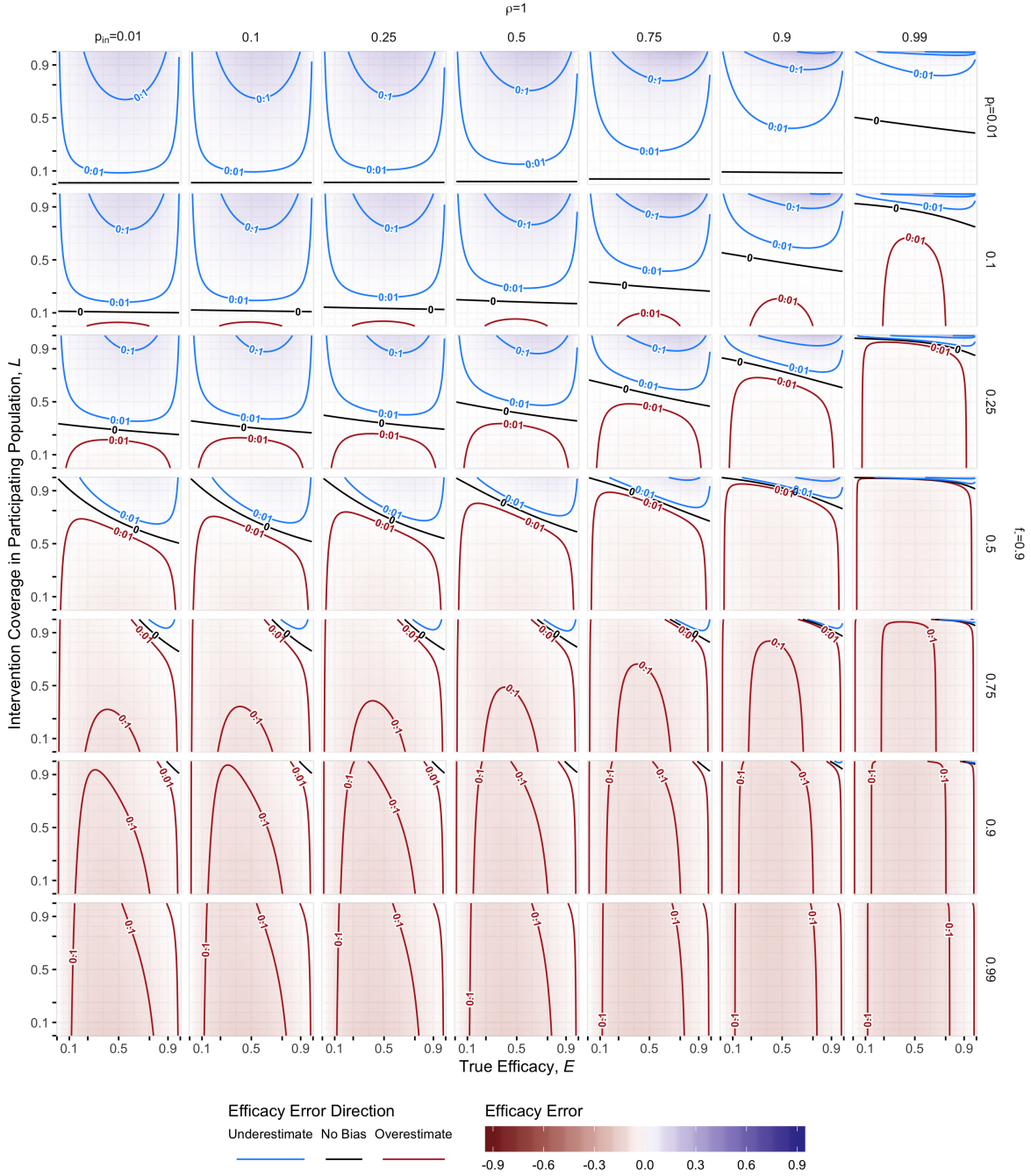


Figure S19: **Bias Sensitivity Without Secondary Test-Positives.**: Recall, $\rho = \frac{\lambda}{B+1}$ is the expected ratio of secondary to primary recruits; $p_t = R''/\lambda$ is the expected fraction of secondary recruits that test positive when no intervention is present; and $f_- = \frac{B}{B+1}$ is the expected fraction of primary recruits that are test-negative. *In this panel, $\rho = 1$ and $f_- = 0.9$.*

270 S6 Calculation of Coverage, L , and Participation, p_{in}

271 In addition to parameters associated with epidemiology and response activities (B , λ , and R''), the values of L
272 and p_{in} are required to determine what level of bias may be present. Our model proposes a population where
273 there are three groups, non-participants, and participants that receive the intervention or not. If we have
274 an estimate for the size of the total recruitable population, we can use measures taken during the intervention
275 distribution to estimate these values. Alternatively, additional data could be collected when testing recruits
276 to estimate these values. The estimates proposed hereafter are not definitive; in addition to assuming our
277 model of heterogeneity is a sufficiently useful approximation, they make further strong assumptions about
278 behaviour around intervention uptake. However, as estimates they are potentially informative about the
279 limits of p_{in} and L within the model framework we have proposed.

280 S6.1 Measures During Intervention Distribution

281 During distribution of the intervention, one could count the number of people receiving the intervention
282 (V) and the number ineligible (U^*). From those values we can compute the crude minimum and maximum
283 values of L and p_{in} (where $\max(L)$ corresponds to $\min(p_{\text{in}})$, and *vice versa*).

$$\begin{aligned} L &\in \left(\frac{V}{T}, \frac{V}{V + U^*} \right) \\ p_{\text{in}} &\in \left(0, 1 - \frac{V + U^*}{T} \right) \end{aligned} \quad (\text{S33})$$

284 In the model, we make no assumptions of how non-intervention occurs in the participating population,
285 just that it occurs randomly within that group. Thus the ineligible count represents the minimum non-
286 intervention amongst the participating population (the upper limit of L); there may be other sources (*e.g.*,
287 participants are unavailable on the day offered). Potentially, when distributing the intervention, participants
288 could be asked about members of their household, neighbors, *etc.* that wanted to get the intervention,
289 but were unable to do so, but this number would also have many uncertainties (*e.g.* duplicate reporting,
290 reporting individuals that do receive the intervention at a different time or place).

291 If the study intervention has multiple steps (*e.g.* a two-dose vaccine, repeat application of vector-control
292 insecticides), then the decrease in coverage between steps could be informative about the participation rate,
293 p_{in} . If we assume not receiving the intervention is due to a mix of short-term (*e.g.* ill that day) and long-term
294 (*e.g.* too young to be eligible) effects, then we can potentially further constrain L and p_{in} . In the following
295 we assume that: i) long-term ineligibles only present themselves at the first step (though they may also
296 not), ii) short-term ineligibles present at the same rate in the subsequent steps, and iii) individuals that
297 did not present at earlier steps will also not present later. If we apply these assumptions to a two-step
298 intervention, and we call unobserved long-term ineligibles I_0 , the long-term ineligibles that appear initially
299 are I_1 , and the intervention recipients (V) and short-term ineligibles that present (U^*) or not (U) at each
300 stage ($V_1, V_2, U^*_1, U^*_2, U_1, U_2$), then the following relations hold.

301 We have six observed pieces of information: the total population (T), the number given the intervention
302 versus short term ineligible at both steps (V_1, V_2, U^*_1, U^*_2), and the number of long-term ineligibles at the
303 first step (I_1).

304 We know that at the second intervention step, we have only people that got vaccinated in the previous step,
305 no new long term ineligibles, and the breakdown of short term ineligibles versus those that receive the second
306 step. So:

$$V_1 = U_2 + U_2^* + V_2 \implies U_2 = V_1 - V_2 - U_2^* \quad (\text{S34})$$

307 If we assume that the division between presenting versus not presenting for short-term ineligible is the same
 308 for both steps, then:

$$\begin{aligned} \frac{U_2^*}{U_2 + U_2^*} = \frac{U_1^*}{U_1 + U_1^*} &\implies \frac{U_2^*}{U_2 + U_2^*} (U_1 + U_1^*) = U_1^* \\ &\implies \frac{U_2^*}{U_2 + U_2^*} U_1 = U_1^* \left(1 - \frac{U_2^*}{U_2 + U_2^*}\right) \\ &\implies U_1^* \frac{V_1 - V_2 - U_2^*}{U_2^*} = U_1 \end{aligned} \quad (\text{S35})$$

309 For the first distribution of the intervention, we assume that the ratio of short- and long-term ineligible
 310 presenting is the same as those not presenting:

$$\begin{aligned} \frac{U_1^*}{U_1^* + I_1} = \frac{U_1}{U_1 + I_0} &\implies \frac{U_1^*}{U_1^* + I_1} (U_1 + I_0) = U_1 \\ &\implies \frac{U_1^*}{U_1^* + I_1} I_0 = U_1 \left(1 - \frac{U_1^*}{U_1^* + I_1}\right) \\ &\implies I_0 = U_1 \frac{I_1}{U_1^*} = I_1 \frac{V_1 - V_2 - U_2^*}{U_2^*} \end{aligned} \quad (\text{S36})$$

311 Finally, the pieces must add up to the total population, and therefore:

$$\begin{aligned} C &= V_1 + U_1^* + I_1 + U_1 + I_0 \\ &= V_1 + U_1^* \left(1 + \frac{V_1 - V_2 - U_2^*}{U_2^*}\right) + I_1 \left(1 + \frac{V_1 - V_2 - U_2^*}{U_2^*}\right) \\ &= V_1 + \frac{V_1 - V_2}{U_2^*} (U_1^* + I_1) \\ T &= C + N \\ N &= T - C \\ &= T - V_1 - \frac{V_1 - V_2}{U_2^*} (U_1^* + I_1) \end{aligned} \quad (\text{S37})$$

312 So using these assumptions, we can estimate the coverage and the participating fraction:

$$\begin{aligned} L = \frac{V}{C} &= \frac{V_2}{V_1 + \frac{V_1 - V_2}{U_2^*} (U_1^* + I_1)} \\ p_{\text{in}} = \frac{C}{T} &= \frac{V_1 + \frac{V_1 - V_2}{U_2^*} (U_1^* + I_1)}{T} \end{aligned} \quad (\text{S38})$$

313 Both of these equations consist only of the measured values. Since the model is an approximation, there
 314 may be other effects, but these relations can provide a useful guide to the value of the study parameters that
 315 contribute to bias.

316 S7 Alternative Scenario Translation

317 In the main text, we described applying this model to evaluating a novel vaccine during an Ebola outbreak,
318 but noted in the discussion that the approach could be generically applicable. The previous sections outline
319 the model in generic terms. Here we provide an example translation of that generalisation to another case:
320 a vector control intervention for dengue.

321 In this scenario, we consider an intervention like indoor residual spraying, applied to urban households on
322 a block basis (*i.e.* set of contiguous households, determined by street intersections) ahead of the dengue
323 season. Some blocks would get no coverage (*i.e.* be amongst the non-participating population), while others
324 would receive coverage at some level (with non-coverage corresponding to *e.g.* availability to let treatment
325 teams into house on that day or presence of children under some age).

326 Later, during the dengue season, people in the study population would seek healthcare with symptoms that
327 would lead to testing for dengue, corresponding to the primary process. However, because dengue is fre-
328 quently asymptomatic, the secondary process would be to test individuals in the primary cases household and
329 adjacent households. Instead of a contacts-based secondary route, there is geospatial secondary route.

330 Whether a TNCC study would be ideal for this scenario is certainly a topic for debate. However, it is possible
331 to frame this scenario and other potential pathogen spread and surveillance processes in the same terms we
332 have introduced in this analysis.