# Residual infestation and recolonization in an urban *Triatoma infestans* control campaign Supplementary Materials

C. M. Barbu et al.

## Contents

# 1 Modalities of insecticide treatment in initial treatment phase

The treatment phase used pyrethroid insecticides, mainly deltamethrin (96.4%), in 5 formulations over the 9 years of the treatment covered here (Tab. S1). The formulation was diluted in water and applied, using a Hudson X-Pert compression sprayer at a target dose rate of 30 mg/m$^2$ for deltamethrin based insecticides and 36 mg/m$^2$ for lambda-cyhalothrin based insecticides.

| Formulation | Years | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | |
| DELTA 5% SC | 0 | 0 | 0 | 5,323 | 10,399 | 13,095 | 9,786 | 3,286 | 41,889 |
| K-OTHRINE 50 SC | 3,977 | 4,062 | 1,005 | 1,416 | 2 | 0 | 0 | 0 | 10,462 |
| K-OTHRINE 5% PM | 0 | 8,982 | 11,094 | 8,689 | 0 | 0 | 0 | 0 | 28,765 |
| LAMBDA 10% PM | 0 | 27 | 0 | 0 | 3,001 | 0 | 0 | 0 | 2,646 |

Table S1: Number of households treated with each insecticide formulation in treatment phase per year DELTA 5% SC: suspension concentrate of deltamethrin. K-OTHRINE 50 SC: suspension concentrate of deltamethrin. K-OTHRINE 5% PM: wettable powder of deltamethrin. LAMBDA 10% PM: wettable powder of lambda-cyhalothrin. DEMAND 10 CS: capsule suspension of lambda-cyhalothrin.

# 2 Model of observation and infestation during the treatment phase

## 2.1 General model

The simple model of observation and infestation that we developed links the observed infestation in households that were treated twice to the real infestation prevalence accounting for $s$, the sensitivity of the inspectors, and $c$, the probability of clearing a household from infestation with one treatment.

We focused on households treated twice to evaluate the effectiveness $c$ of the treatment, as each treatment also corresponds to the observation of the infestation. Among all the households treated twice:

- $O_{I+II+}$ were observed infested twice
- $O_{I+II-}$ were initially observed infested then observed uninfested
- $O_{I-II+}$ were initially observed uninfested but then observed infested
- $O_{I-II-}$ were observed non-infested twice

We assumed that households did not become infested between the two treatments, for the following reasons: the treatments were separated by only six months; the overall infestation is severely reduced by the first treatment; and treated households are normally protected by the treatment for more than three months [?, ?].

## 2.2 Deterministic version of the model

The real initial number of infested households $n_{I/II}$, observed and treated twice, is then observed according to $s$ the sensitivity of the inspectors and $c$ the probability of elimination of the infestation by the first treatment according to:

$$\left\{\begin{array}{lcl} O_{I+II+} & = & n_{I/II} \cdot s \cdot (1-c) \cdot s \\ O_{I+II-} & = & n_{I/II} \cdot s \cdot (c + (1-c)(1-s)) \\ O_{I-II+} & = & n_{I/II} \cdot (1-s) \cdot (1-c) \cdot s \end{array}\right. \tag{1}$$

which can be solved algebraically:

$$\left\{\begin{array}{rcl} n_{I/II} & = & \frac{0_{I+II+}}{s^2 \cdot (1-c)} \\ c & = & 1 - \frac{1}{s \cdot \left(1 + \frac{0_{I+II-}}{0_{I+II+}}\right)} \\ s & = & \frac{1}{1 + \frac{0_{I-II+}}{0_{I+II+}}} \end{array}\right. \tag{2}$$

## 2.3 Stochastic version of the model

To assess the robustness of our results to stochastic variability, the model in eq.1 can be written as the realization of successive binomial events (Fig. S1) when separately considering $O_{I+}$, the number of households observed infested during the first observation/treatment.

As these three events are independent, the overall likelihood of the model is given by:

$$ll\left(n_{I/II}, c, s\right) = P(O_{I+}) \cdot P(O_{I+II+}|O_{I+}) \cdot P\left(O_{I-II+}|O_{I+}\right) \tag{3}$$

which can be sampled easily by Monte Carlo Markov Chain (MCMC), assuming a flat prior for all parameters. After 100,000 iterations the chains were well converged, and we obtained smooth distributions of $c$,$s$ and $n_{I/II}$ (Fig. S2).

The MCMC and the analytical solutions are in near perfect agreement. The distribution of $c$, the effectiveness of the treatment, is extremely narrow (0.987 [0.984-0.989]) but $s$, the sensitivity of the inspectors, is quite broad (0.56 [0.46-0.66]) leading to a large distribution of $n_{I/II}$, the initial number of infested households (12,737 [10,771-15,407]).

The other equations in table 2 of the main text also allow us to compute the number of residually infested households given the model parameter values. We computed the residual infestation for each iteration of the MCMC and generated the distribution and credible interval at 95% of the residual population size for each category of participation (Table S2). Even when accounting for stochastic variability in the data, the estimated share of non-participating households in residual populations remains in a narrow range between 96.4% and 97.6%.

| First observation | | Second observation |
|---|---|---|
| $O_{I+} \sim B\left(n_{I/II}, s\right)$ | + | $O_{I+II+} \mid O_{I+} \sim B\left(O_{I+}, (1-c)\cdot s\right)$ |
| | - | $O_{I-II+} \mid O_{I+} \sim B\left(n_{I/II} - O_{I+}, (1-c)\cdot s\right)$ |

Figure S1: **Stochastic model of infestation observation during the treatment phase in households observed and treated twice** + and - stand for second observation of households respectively infested and non-infested at first observation. The $n_{I/II}$ infested households have a probability $s$ of being observed infested at the first observation resulting in $O_{I+}$ households observed positive and $O_{I-} = n_{I/II} - O_{I+}$ infested households observed non-infested. The households observed infested have then a probability $(1-c)$ to remain infested after treatment and an independent probability $s$ to be observed positive again, resulting in $O_{I+II+}$ households observed twice. Independently, the households initially infested but observed negative also have a probability $(1-c)\cdot s$ to be observed infested at the second observation.

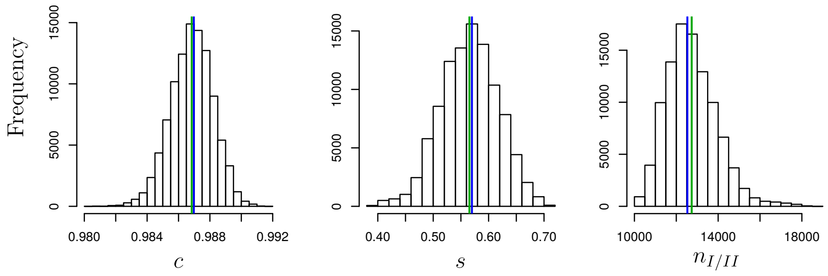

Figure S2: **Posteriors of treatment phase model parameters.**
The green vertical line corresponds to the mean of the sampled values; the blue vertical line corresponds to the analytical solution

## 2.4 Impact of considering the observation to be less than perfect

The imperfect observation of the infestation as been repeatedly suggested, both during surveys [1] and treatment [2]. We point out here that our assumption that the observation of the infestation is not perfect is a conservative one and show that most of the infestation post-treatment is due to non-participating households.

Table S2: **Number of residually infested households as estimated by the stochastic model, with 2.5% and 97.5% quantiles (95% credible interval).**

| Treatment | Mean | 2.5% | 97.5% |
|---|---|---|---|
| I & II | 2.2 | 1.4 | 3.4 |
| Ionly | 14.6 | 11.0 | 19.1 |
| IIonly | 3.9 | 2.9 | 5.1 |
| None | 677.5 | 572.1 | 821.1 |
| Total | 698.1 | 589.5 | 846.0 |
| Share None | 97.0% | 96.4% | 97.6% |

If the sensitivity is perfect ($s = 1$), the system of equations 1 simplifies to:

$$\begin{cases} O_{I+II+} &= n_{I/II} \cdot (1-c) \\ O_{I+II-} &= n_{I/II} \cdot c \end{cases} \tag{4}$$

$$\Leftrightarrow \begin{cases} n_{I/II} &= \frac{O_{I+II-}}{c} \\ O_{I+II+} &= \frac{O_{I+II-}}{c} * (1-c) \end{cases} \tag{5}$$

$$\Leftrightarrow \begin{cases} n_{I/II} &= \frac{O_{I+II-}}{c} \\ c &= \frac{1}{1+\frac{O_{I+II+}}{O_{I+II-}}} \end{cases} \tag{6}$$

When assuming perfect observation, $c = 1/(1+53/7087) = 99.3\%$ compared to 98.7%, the estimate when allowing for imperfect observation. The difference might seem slight, but the estimated fraction of infestation remaining after one treatment is reduced by 43% and after two treatments by 66% from the increase in treatment effectiveness, $c$. This significantly higher estimate of the effectiveness of the treatment implies that an even larger proportion of residual infestation results from non-treated households since treated households now have a greater probability of clearing the infestation. By incorporating the imperfect quality of the inspectors, we are being conservative regarding our main hypothesis.

## 2.5 Sensitivity analysis

### 2.5.1 Sensitivity to independence of the two treatments

In the main text (Table 3), we consider the effectiveness of the two treatments to be equal and independent. However, if structural reasons limit the effectiveness of the first treatment in a specific household, it is likely that these same limitations apply again to the second treatment.

To estimate the sensitivity of our results to this alternative hypothesis, we consider the second treatment to be completely ineffective.

Table S3: **Sensitivity of the treatment phase effectiveness model to variations of the hypotheses**

| Treatment Phase Treatment | Main Model | Residual population<br>No effect $2^{nd}$ treatment | Non-treated less infested |
|---|---|---|---|
| I & II | 2 (0.3%) | 163 (19.3%) | 2 (1.4%) |
| I only | 14 (2.1%) | 14 (1.7%) | 14 (9.2%) |
| II only | 4 (0.5%) | 4 (0.4%) | 4 (2.5%) |
| None | 666 (97.1%) | 666 (78.6%) | 133 (86.9%) |
| Total | 686 (100%,0%) | 847 (100%,+23%) | 153 (100%,-78%) |

Main model corresponds to the deterministic model described above. No effect $2^{nd}$ treatment corresponds to applying the treatment removal $c$ only once to infested households treated twice - that is, the second treatment has no effect. Non-treated less infested corresponds to the extreme case where non-treated households are initially 5 times less infested than the households treated only at second treatment ($p_{\varnothing} = p_{II}/5$). Estimates are given in rounded number of households. Between parentheses is the proportion of residually infested households in each category of participation for each model. On the last line, the second percentage in parentheses is the percent increase compared to the main model.

Even in this extreme case, non-participating households represented most of the households still infested after the treatment phase. In addition, the overall number of households residually infested after the treatment phase only increased by 24% (Table S3, third column).

In practice, none of the households observed infested in both treatments of the treatment phase were found infested during surveillance inspections, suggesting noticeable effectiveness of the second treatment.

### 2.5.2 Sensitivity to the correlation of infestation and participation

When estimating the residual infestation, we used the infestation prevalence among households that participated only in the second treatment of the treatment phase as a proxy of the infestation prevalence in households that never participated

However, we observe that households participating in only one treatment have a significantly lower prevalence of infestation (Table 4) suggesting the prevalence in households not participating in any treatment may be even lower. As the prevalence in households treated only the second time is 5 times lower than in households participating in both treatments, we consider the extreme case where the prevalence in non-participating households is 5 times lower than in households participating only in the second treatment.

Even this strong modification does not qualitatively change our results: over 85% of the residual infestation is still in non-participating households, and the overall effectiveness of the treatment phase would be even higher with a decrease of 78% of the number of households presenting some residual infestation after the treatment phase (Table S3, fourth column).

The limited number of reports (225 on 164 city blocks) during the surveillance phase suggests that the residual infestation may be much closer to this last "extreme" estimate than to our initial hypothesis. The infestation in non-participating households may then be much lower than in households treated only once. Alternatively, or simultaneously, active dispersal between non-treated households and treated households may generate some protection for non-treated households as insects in non-treated households may be eliminated during migrations. An important dispersal dynamic would also help explain the very low rates of detection of the infestation by our trained inspectors compared to the inhabitants: the inhabitants would detect transient infestation much better than trained inspectors present only for an hour.

# 3 Difference in difference analysis using pre-treatment survey data

Infestation data collected during pre-treatment surveys are currently only available for the district of Mariano Melgar, a very urban, fairly central district. For this district we compared the infestation observed during pre-treatment surveys and during treatment to assess the effectiveness of the treatment.

## 3.1 Data

Surveys were conducted between June 2008 and February 2009. Of 12,870 households in 37 localities, 7,959 households accepted to have their domestic and peridomestic areas to be inspected for nymph's stages and adults, 608 (7.6%) of the inspected households were found infested (at least one $T. infestans$ of any stage but egg found).

Given the results of the survey, 16 localities encompassing 9,801 households were targeted for two treatments between May 2011 and July 2012. 4,416 households were inspected both during the survey and the second treatment; of which 3,512 also received the first treatment (Table S4). Among households observed during the survey, 10 were found infested during the second spray; only one household was observed infested in both treatments.

## 3.2 Analysis and results

### 3.2.1 Significance of the effect of the campaign

We analyze the success of the campaign in this district using a difference-in-difference approach [3, 4]. Focusing on households observed both during the pre-treatment survey and during the second treatment. The outcome we model is the observation of infestation. As the outcome is binary, the infestation in a

| Infested at survey | First treatment | Infested at second treatment | | Total |
|---|---|---|---|---|
| | | FALSE | TRUE | |
| FALSE | FALSE | 859 | 4 | 863 |
| FALSE | TRUE | 3,058 | 1 | 3,059 |
| TRUE | FALSE | 38 | 3 | 41 |
| TRUE | TRUE | 451 | 2 | 453 |
| | Total | 4,406 | 10 | 4,416 |

Table S4: Number of households observed infested during survey and second treatment depending on the realisation of the first treatment

household is modeled with a logistic model in which the treatment in the first wave is included as a covariate:

$$log\left(\frac{p_i}{1-p_i}\right) = \beta_i + \beta_T T + \beta_t t + \beta_{tT} T * t + \epsilon$$

where $T$ is 1 if the household is in the treated group and 0 if not. $t$ is the time of observation: 0 for the pre-treatment survey observation and 1 for the second treatment observation. $\beta_x$ corresponds to the regression coefficient for $x$.

We obtain:

```
Parameter                  Estimate Std. Error  z value Signifs
Intercept (beta_i)            -3.05       0.16   -19.06  ***
Observation time (beta_t)     -1.81       0.41    -4.39  ***
Treated group (beta_T)         1.14       0.17     6.78  ***
Treatment effect (beta_tT)    -3.35       0.71    -4.71  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

N: 4,416 x 2 observations, 494+10 observations of infestation
Nagelkerke pseudo $R^2$ index: 0.22 [?].

Table S5: Difference in difference analysis, regression coefficients

With corresponding odds ratios of:

The effect of the treatment $\beta_{tT}$ is extremely strong and significant (OR: 0.05 [0.03-0.06]), and the correlation between observed infestation and receiving treatment, quantified by the “treated group” effect (T=1), is also strong and significant (OR: 3.1 [2.2-4.3]).

Finally, it is striking that, independently of the treatment, the observed infestation significantly decreases between the survey and the treatment, suggesting, as observed in the analyses of the main text, that the observation of

```
Parameter                      OR   2.5%  97.5%  Signifs
Intercept (beta_i)          0.048  0.035  0.065  ***
Observation time (beta_t)   0.164  0.073  0.368  ***
Treated group (beta_T)      3.117  2.244  4.329  ***
Treatment effect (beta_tT)  0.035  0.009  0.142  ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Table S6: Difference in difference analysis, odds ratios

infestation during the treatment is far from perfect. It is also possible that the decrease is partially due to treatment of the households performed by inhabitants independently of the public health authorities in the 2 to 3 year interval between the survey and the official treatment campaign.

### 3.2.2 Estimation of residual infestation post-treatment

To estimate the residual infestation post-treatment in the difference-in-difference framework, we use the results of the difference-in-difference model and extrapolate them to households not observed during the pre-treatment survey or the second phase of treatment. To guide this extrapolation, we study the distribution of the observed infestation in the pre-treatment survey and the two phases of the treatment (Table S7).

Table S7: **Observed infestation in Mariano Melgar in pre-treatment survey and two phases of the treatment**

| Survey | TP I | TP II | Number of households |
|---|---|---|---|
| + | + | + | 1 |
| | | - | 159 |
| | | ∅ | 13 |
| | - | + | 1 |
| | | - | 292 |
| | | ∅ | 40 |
| | ∅ | + | 3 |
| | | - | 38 |
| | | ∅ | 59 |
| | Not targeted | | 2 |
| - | + | + | 0 |
| | | - | 144 |
| | | ∅ | 18 |
| | - | + | 1 |
| | | - | 2,914 |
| | | ∅ | 817 |
| | ∅ | + | 4 |
| | | - | 859 |
| | | ∅ | 1,177 |
| | Not targeted | | 1,417 |
| ∅ | + | + | 1 |
| | | - | 103 |
| | | ∅ | 15 |
| | - | + | 0 |
| | | - | 1,230 |
| | | ∅ | 583 |
| | ∅ | + | 1 |
| | | - | 552 |
| | | ∅ | 776 |
| | Not targeted | | 1,650 |
| | | Total | 12,870 |

TP I : first spray of the initial treatment phase (attack phase 1). TP II : second spray of the initial treatment phase (attack phase 2). Household infestation: +, treated or surveyed and observed infested; -, treated or surveyed and observed non-infested; ∅, no survey or treatment and infestation not observed.

First, we observe that the proportion of infested households during the pre-treatment survey who chose not to participate in the first and second treatment (4.8%) is close to the percentage of households only participating in the second treatment (4.5%), somewhat lower than households only participating in the first treatment (6.0%), and much lower than households participating in both phases (12.9%). Interestingly, we also find a significant link between participation in the survey and observed infestation during the treatment phase (OR: 1.36 [1.06-1.77], simple fisher test), suggesting that infested households are more likely to participate, even in pre-treatment surveys lacking insecticide application. From these observations, we can use the estimates of the difference-in-difference model for Mariano Melgar to estimate the proportion of residually infested households among treated and non-treated households.

We consider the effect of the treatment to be the same and independent for the first and second phase by applying the interaction term $\beta_{tT}$ as many times as the household received treatment. As previously remarked, we also assume that the upward bias in participation of infested households, $\beta_T$, only applies for households treated twice.

We define the probability of infestation after treatment for all targeted households as:

$$\begin{aligned}
p_2 &= \text{inverse-logit}\left(\beta_i + \beta_T + \beta_t + 2 * \beta_{tT}\right) \\
p_1 &= \text{inverse-logit}\left(\beta_i + \beta_t + \beta_{tT}\right) \\
p_0 &= \text{inverse-logit}\left(\beta_i + \beta_t\right)
\end{aligned}$$

with $p_n$ the probability to be infested after the treatment phase when receiving $n$ treatments.

We estimate that households having received no treatment account for 94.4 % of the infested households after the treatment phase. We obtain very similar estimates running the treatment phase model presented in section 1 on the Mariano Melgar district: 93.6 % [86.9% - 98.0%].

# 4 Estimating the reliability of surveillance

We have reasons to believe that the specificity of the reports is excellent, as insects were systematically collected and identified by the trained personnel of the health center and subsequently by our research team. Quantifying the sensitivity of the reports is much more complex.

To assess the sensitivity of surveillance, we sought to estimate the prevalence of infestation in households that had not reported and thus were not part of surveillance. We carried out an active search in a subsample of these households in 2013 for the districts of Paucarpata, Sachaca, Tiabaya, and Socabaya which were part of the surveillance phase. The subsample was stratified into three arms: A) households within 50 meters of a reporting household that had not been already inspected by the surveillance program, B) randomly selected households in treated areas, and C) non-treated households within 50 meters of a household that was observed infested during treatment. Arm B served as the control arm. A total of 1,676 households were selected to be part of this subsample with 262 households in arm A, 751 households in B, and 663 households in C. Of these households, a total of 740 (44.1%) did not participate. Of the 966 households participating that were searched, we detected only 3 (0.31%) infestations, two in arm A (2/193 or 1.04%) and one in arm C (1/350 or 0.285%) with no positives in the control arm B. Detailed results are given below in Table S8. These results suggest that, while the sensitivity of the passive reporting system is not perfect, it is fairly high.

Table S8: **Results of active search inspections in non-reporting households**

| Arm | + | - | ∅ | Total |
|---|---|---|---|---|
| A | 2 | 191 | 69 | 262 |
| B | 0 | 393 | 358 | 751 |
| C | 1 | 349 | 313 | 663 |
| Total | 3 | 933 | 740 | 1,676 |

Household infestation: ∅ corresponds to not participating; + to inspected positive; - to inspected negative. A total of 1,676 households were surveyed in the active search. Arm A corresponds to households within 50 meters of a reporting household that had not been previously inspected; B corresponds to randomly selected households in treated areas; and, C corresponds to households non-treated during the initial treatment phase and within 50 meters of a household observed infested during the initial treatments.

# 5 Details of observed infestation during surveillance inspections

## 5.1 Infestation history

The table below gives the details of the results of surveillance and treatment phase inspections for all households targeted for treatment (Table S9).

A summary of the surveillance phase in terms of treatment history is also given below in figure S3.

Table S9: **Observed infestation during surveillance according to treatment phase year and status**

| Surv. | TP II | TP I | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2011 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| + | + | + | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | | - | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| | | ∅ | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| | - | + | 7 | 17 | 3 | 0 | 1 | 0 | 2 | 30 |
| | | - | 6 | 25 | 1 | 3 | 1 | 1 | 1 | 38 |
| | | ∅ | 4 | 2 | 0 | 0 | 2 | 0 | 0 | 8 |
| | ∅ | + | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 5 |
| | | - | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 3 |
| | | ∅ | 1 | 6 | 2 | 4 | 9 | 3 | 4 | 29 |
| - | + | + | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| | | - | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| | | ∅ | 0 | 2 | 1 | 0 | 1 | 2 | 0 | 6 |
| | - | + | 18 | 51 | 18 | 6 | 18 | 2 | 1 | 114 |
| | | - | 54 | 98 | 37 | 21 | 31 | 7 | 7 | 255 |
| | | ∅ | 3 | 6 | 2 | 1 | 10 | 2 | 2 | 26 |
| | ∅ | + | 2 | 3 | 1 | 0 | 1 | 0 | 0 | 7 |
| | | - | 3 | 6 | 3 | 4 | 4 | 1 | 7 | 28 |
| | | ∅ | 3 | 7 | 6 | 11 | 19 | 5 | 6 | 57 |
| ∅ | + | + | 10 | 16 | 10 | 4 | 3 | 6 | 2 | 51 |
| | | - | 5 | 14 | 9 | 5 | 0 | 0 | 4 | 37 |
| | | ∅ | 16 | 21 | 27 | 40 | 19 | 17 | 17 | 157 |
| | - | + | 781 | 1,218 | 1,633 | 1,395 | 1,031 | 530 | 355 | 6,943 |
| | | - | 2,943 | 5,472 | 4,916 | 5,169 | 3,204 | 3,757 | 2,273 | 27,734 |
| | | ∅ | 432 | 602 | 597 | 716 | 518 | 511 | 594 | 3,970 |
| | ∅ | + | 35 | 71 | 69 | 73 | 168 | 97 | 95 | 608 |
| | | - | 275 | 836 | 627 | 839 | 1,514 | 1,331 | 1,448 | 6,870 |
| | | ∅ | 257 | 1,035 | 831 | 2,034 | 1,485 | 1,913 | 1,953 | 9,508 |
| | Total | | 4,856 | 9,519 | 8,796 | 10,325 | 8,039 | 8,185 | 6,771 | 56,491 |

Household infestation: +, treated, observed infested; -, treated, observed non-infested; ∅, no treatment and infestation not observed. The year is the year of last action of treatment phase in the locality. TP I : first spray of the initial treatment phase (attack phase 1). TP II : second spray of the initial treatment phase (attack phase 2).

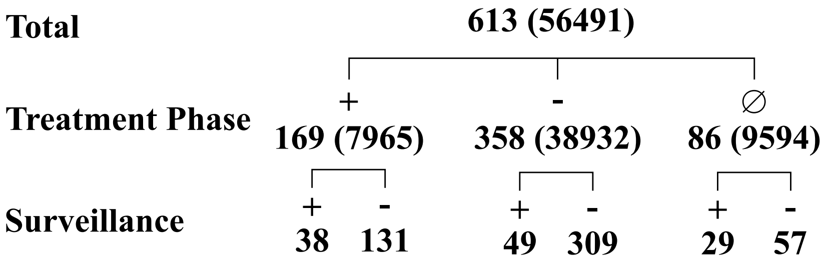

Figure S3: **History of insecticide application and infestation in households inspected during the surveillance phase in Arequipa, Peru from 2009 to 2012 after treatment of the localities in treatment phase between 2004 and 2011.**

Household infestation: ∅ , no treatment and infestation could not be observed; + , treated at least once and observed at least once infested; - , treated at least once and never observed infested. The first number corresponds to the number of households observed in surveillance for this outcome of the treatment phase. The numbers between parenthesis correspond to the total number of households in our sample for this outcome of the treatment phase.

## 5.2 *T. infestans* population structure

Complete data on the insects collected during the surveillance phase are available below. Nymph stages predominate (80%); among the nymphs smaller stages are under represented (fig. S4) compared to classical population structure for *T. infestans* [5]. The underrepresentation of early-stage nymphs is expected, as smaller juveniles are more difficult to detect and collect.

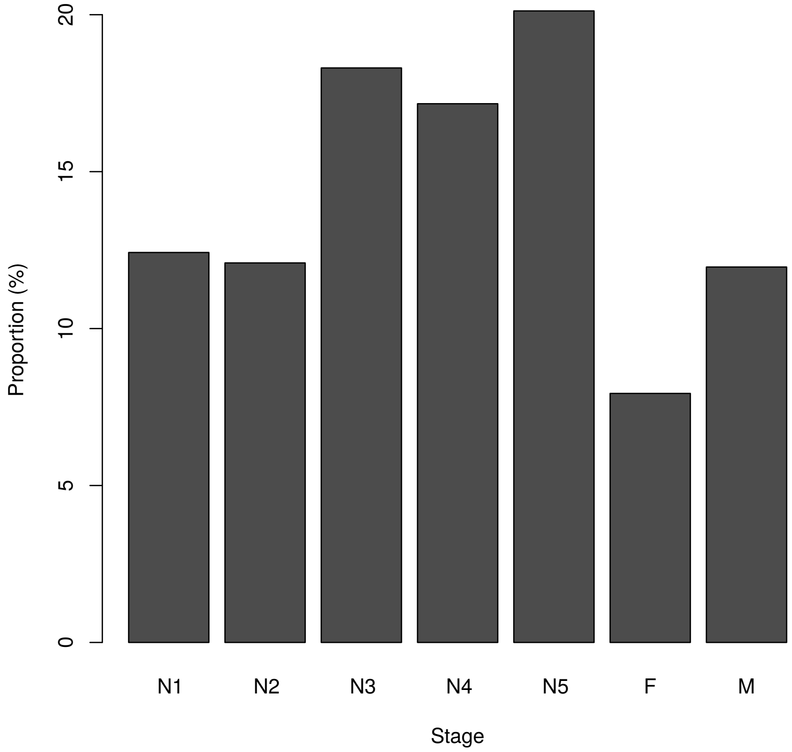

Figure S4: **Average observed population structure of *T. infestans* in households treated in the surveillance phase.**
N1-5: nymphs 1-5, F: females, M: males. Eggs were not collected.

The total population size has a median of 8 insects, a mean of 40.7, and a standard deviation of 102.4. The full distribution of the population sizes is presented in figure S5.

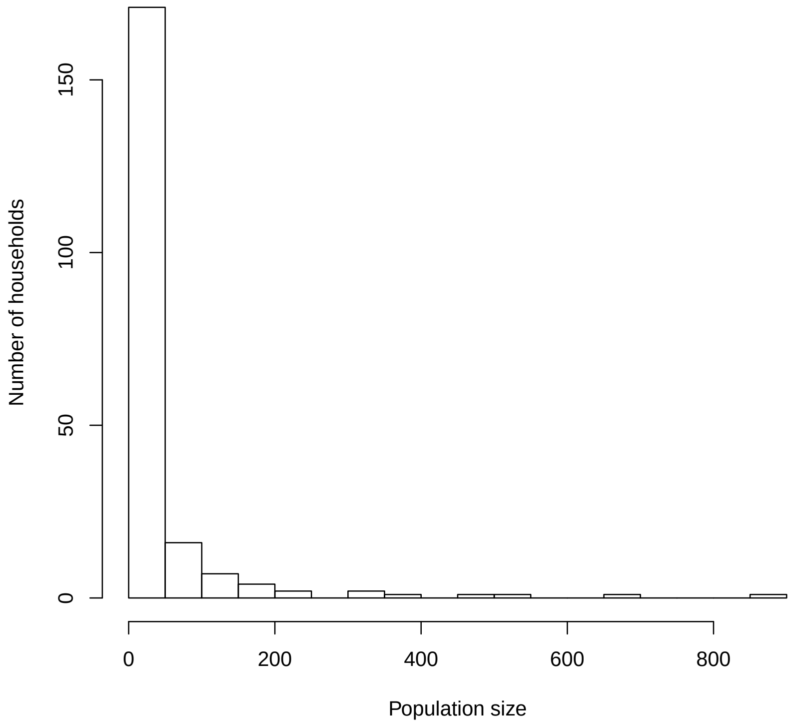

Figure S5: **Distribution of *T. infestans* population sizes in households treated in the surveillance phase.**

## 6 Spatial autocorrelation of the infection by *T. cruzi* of *T. infestans* among infested households

To assess if houses infected by *T. cruzi* are spatially clustered we calculated the correlogram of the infection among infested households and the significance of the correlation using 1,000 permutations. We performed this analysis using the package $ncf$ in R.

Once applying the Bonferroni correction, only spatial correlation within 50 meters is significant (fig. S6).

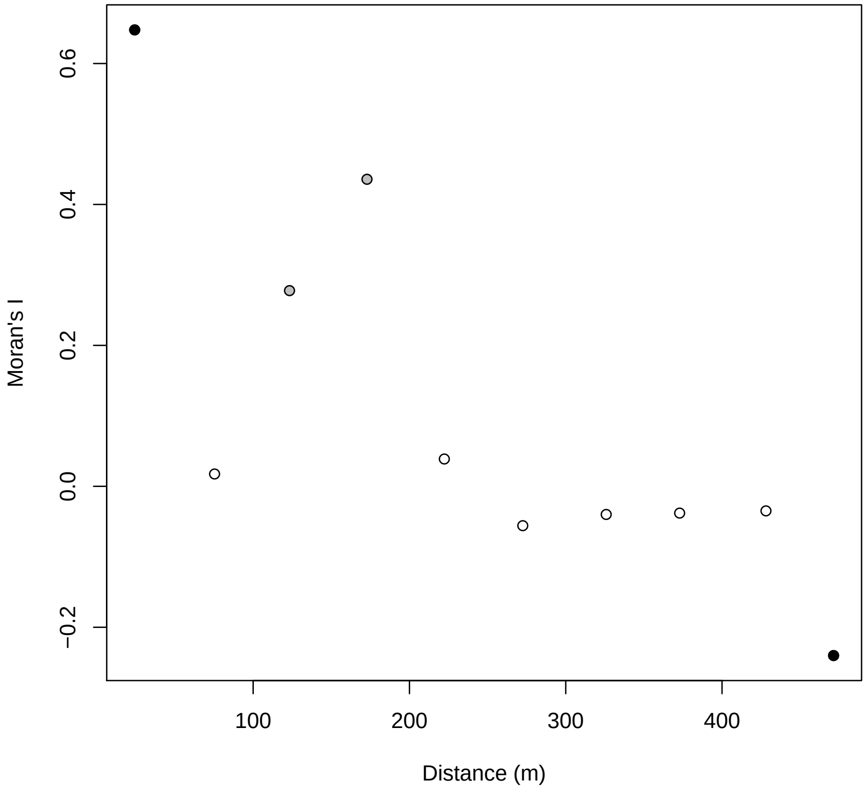

Figure S6: Correlogram of the presence of *T. cruzi* in households infested by *T. infestans* The spatial autocorrelation is estimated applying the Moran's I to distance classes of 50 meters. Grey points correspond to Moran's I values significantly different from the expectation under the null hypothesis using $\alpha = 0.05$, black points correspond a significance threshold of $\alpha/10$ corresponding to the multiple testing Bonferroni correction.

## 7 Models of infestation in surveillance: model selection

We considered the following cofactors:

- non-participation to the treatment phase (treated $= 0$ in the model)
- infestation observed in at least one treatment (inf $> 0$ in the model)
- time in years between the end of the treatment phase and December 2012

As the observation of infestation and non-participation are mutually exclusive, we consider all the factors and interactions that can be considered:

- Not treated
- Infestation
- Time post-intervention (Time PI)
- Time PI : Treated (interaction between time post-intervention and non-treatment)
- Time PI : Infestation (interaction between time post-intervention and initial infestation)

resulting in 31 possible combinations. Some of them are equivalent because interactions including the time post-intervention (an integer) imply consideration of the time alone, resulting in a total of 19 uniquely different models.

We calculated the Akaike Information Criterion (AIC) [6] for all 19 different models and ranked them. Hereafter, we present the results for the four models within 3 AIC ranks of the best model presented in the main text.

Table S10: **Best models of infestation in surveillance**

| AIC | Parameters | OR | 2.5% | 97.5% | Signif. |
|---|---|---|---|---|---|
| 1,417.8 | (Intercept) | 5.73e-05 | 1.18e-05 | 2.78e-04 | *** |
| | Not treated | 3.55e+00 | 2.04e+00 | 6.17e+00 | *** |
| | Time post intervention | 1.30e+00 | 1.01e+00 | 1.69e+00 | * |
| | Infestation:Time | 1.20e+00 | 1.11e+00 | 1.29e+00 | *** |
| 1,417.5 | (Intercept) | 2.24e-05 | 2.76e-06 | 1.82e-04 | *** |
| | Not treated | 2.02e+01 | 2.58e+00 | 1.58e+02 | ** |
| | Infested | 2.64e+00 | 1.58e-01 | 4.39e+01 | |
| | Time PI | 1.53e+00 | 1.11e+00 | 2.13e+00 | * |
| | Infested:Time PI | 1.03e+00 | 6.81e-01 | 1.56e+00 | |
| | Not treated:Time PI | 7.35e-01 | 5.23e-01 | 1.03e+00 | . |
| 1,416.0 | (Intercept) | 2.91e-05 | 4.45e-06 | 1.90e-04 | *** |
| | Not treated | 1.56e+01 | 2.49e+00 | 9.84e+01 | ** |
| | Time PI | 1.48e+00 | 1.09e+00 | 1.99e+00 | * |
| | Infested:Time PI | 1.19e+00 | 1.10e+00 | 1.28e+00 | *** |
| | Not treated:Time PI | 7.62e-01 | 5.57e-01 | 1.04e+00 | . |
| 1,415.5 | (Intercept) | 2.11e-05 | 3.16e-06 | 1.41e-04 | *** |
| | Not treated | 2.16e+01 | 3.35e+00 | 1.38e+02 | ** |
| | Infested | 3.21e+00 | 1.96e+00 | 5.24e+00 | *** |
| | Time PI | 1.55e+00 | 1.15e+00 | 2.09e+00 | ** |
| | Not treated:Time PI | 7.28e-01 | 5.32e-01 | 9.95e-01 | * |

AIC: Akaike Information Criterion where lower is better. Parameter: factors in the model. OR: odds-ratio. 2.5% and 97.5%: confidence interval for the parameter value. Signif.: significance code for this factor depending on the p-value: . p<0.1; * p<0.05; ** p<0.01; *** p<0.001.

All these models present a strong effect of the lack of participation (treated equals 0) and a strong effect of time post-intervention. The second best model, with a likelihood very similar to the best model, does not include any effect of the initial infestation but does include a very strong positive interaction between infestation and time post-intervention, which corresponds perfectly to the scenario described by our first analysis of the treatment phase: nearly all the residual infestation after the treatment phase is initially concentrated in non-participating houses (treated equals 0). With time, the probability of infestation increases in all households. The infestation probability increases fastest in previously infested but treated households which indirectly decreases the importance of non-participation as a risk factor.

The best model is similar but with a stable association in time between previous infestation and infestation observed in surveillance phase and an explicit and significant dilution of the non-participation effect with time as recolonization affects all participating households. The preferential recolonization of previously infested households is indirectly present in this model, mediated by the fixed effect of the previous infestation. If persistent infestation could explain this effect just after treatment, the link would progressively be diluted, as we see for non-treated households, and a significant negative interaction with time post-intervention would be selected by the model selection procedure.

In both cases, as recolonization takes place, time tends to decrease the importance of non-participation in the treatment phase and increase the importance of previous infestation on the infestation occurring during the surveillance phase.

## 8 Code to reproduce the methods

The latest version of the code used in this article is available at `http://www.spatcontrol.net/articles/Barbu2014/`.

## References

[1] Hong A, Barbu C, Small D, Levy M (2014) Mapping the spatial distribution of a disease transmitting insect in the presence of surveillance error and missing data. In prep before resubmission to the Journal of the Royal Statistical Society.

[2] Barbu C, Hong A, Manne J, Small D, Calderón J, et al. (2013) The effects of city streets on an urban disease vector. PLOS Computational Biology 9: e1002801.

[3] Card D, Krueger AB (1994) Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. The American Economic Review 84: 772-793.

[4] Branas CC, Cheney RA, MacDonald JM, Tam VW, Jackson TD, et al. (2011) A difference-in-differences analysis of health, safety, and greening vacant urban space. American Journal of Epidemiology 174: 1296–1306.

[5] Rabinovich J (1972) Vital statistics of triatominae (hemiptera: Reduviidae) under laboratory conditions. (i. triatoma infestans klug). Journal of Medical Entomology 9: 351–370.

[6] Sakamoto Y, Akaike H (1978) Analysis of cross classified data by aic. Annals of the Institute of Statistical Mathematics 30: 185–197.