Appendix S2: Derivation of statistical analysis

In the model presented here we assumed that all dispersing female bugs dispersed independently of each other and arrived at and established new colonies at other sites with equal probability. We found no evidence contrary to this assumption in preliminary analysis. We assumed either a high or a low probability of a bug finding other sites, leading to two different scenarios with different link functions and error variances. Since both regression models yielded the same best model we present here only one.

According to this regression model, the probability per bug of arrival at any site was low, making the pool of dispersers much larger than the total number of bugs arriving at any site. Since the effect of bugs arriving at individual sites on the overall pool of dispersers is negligible in this scenario, arrival events on different sites are approximately independent. Furthermore the combination of a large number of dispersers and a low arrival probability leads to a Poisson distribution of arrival events per site (Pitman 1997) with parameter $\lambda = Nr$, where N is the number of dispersers and r is the arrival probability of a disperser on an individual site. Denoting by q the conditional probability of establishment, given arrival, the overall probability of establishment per site becomes:

$$p = 1 - \sum_{k=0}^{\infty} (1-q)^k \frac{\lambda^k e^{-\lambda}}{k!} = 1 - \frac{e^{-\lambda}}{e^{-\lambda(1-q)}} = 1 - e^{-Nqr} .$$

The exponent, $-N*q*r$, was estimated by $a + b*x$, where x indicates the number of bugs found on source sites, leading to the regression equation in the text.

We obtained the maximum likelihood parameter values using iterative weighted least square regressions (McCullagh & Nelder 1989). Our estimation procedure stopped after 30 iterations or when all parameter estimates changed between two iterations by less than 0.1%. We recorded for each model whether or not parameter estimates converged. In rare occasions models yielded an estimate of p below zero; in these cases the estimate was replaced by 10^{-4}.

Classifying a site as infested or uninfested

The survey team did not always survey all existing sites but rather excluded sites it believed to be empty based on extensive prior experience. We carried out two full analyses of the data. In one analysis, reported here, we treated unsurveyed sites as uninfested, as the survey team assumed. In a second, separate analysis, not reported here, we did not assume that unsurveyed sites were uninfested and calculated the explanatory variable as proportion of surveyed sites. That analysis yielded in general similar results but overall a worse model fit, and was therefore considered less informative.

We estimated a separate model based on the assumption that all observed establishment events were a result of misclassifying each season a certain proportion of truly infested sites as uninfested. This model yielded a much worse fit than the models discussed here and is therefore not presented.

Explanatory variable: source of dispersers

The regression model was calculated separately for all combinations of the two alternative response variables described above and 36 alternative explanatory variables. These 36 = 9*2*2 alternatives emerged as follows. We had nine alternative ways of determining source sites at t (9 alternatives). For each alternative we either summed the number of source sites or the number of bugs on source sites at *t* (2 alternatives). We calculated these sums either separately for each ecotype or for all source sites together (2 alternatives). All alternatives for calculating the explanatory variable are listed in Table 1. The nine alternatives to determine which sites infested at *t* could be source sites at *t* arose because the infestation status of a site at each “adjacent” survey, *t* +1 and *t*-1, could influence the site’s status as source site at *t* in three different ways.

The infestation status of a site at survey *t*-1 could affect its contribution to establishment between *t* and *t*+1 if there was a time lag between a site receiving and emitting dispersers such that only sites which were infested at *t and t*-1 emitted dispersers between *t* and *t*+1. Such a time lag could arise in two ways: (i) if only adults dispersed to other sites, it would take a complete bug life cycle between the time when a site received dispersers (i.e. received adults from a different site) and produced its own adult dispersers; (ii) due to local population dynamics, sites could undergo a time lag between receiving dispersers and reaching a density at which bugs dispersed to other sites.

The average time from egg to adult is 161 days (Rabinovich 1972) in optimal laboratory conditions at 27°C and is roughly similar to the time (183 days) between two surveys. If only adults dispersed and the life cycle was somewhat longer in nature, one would not expect sites that were uninfested at *t*-1 and had no nymphs present at *t* to produce dispersers before *t*+1. Hence as a first possibility we defined source sites as sites which were infested at *t* and were either infested at *t*-1 as well or uninfested at *t*-1 but had nymphs present at *t* (Table 1, alternative Ia).

In closed experimental chicken huts, the bug densities of huts initially stocked with five female and three male *T. infestans* continued to increase over two years

(Cecere, Canale & Gürtler 2003), suggesting the possibility of a time lag due to local population dynamics. To describe such a time lag, our second definition of source sites at survey *t* included all sites which were infested at surveys *t and t*-1 (Table 1 alternative I b). The third definition of source sites was simply all infested sites at survey *t*, describing a situation without time lag (Table 1, alternative I c). As mentioned above, we analyzed the nine half-year intervals between surveys starting November 1994. We used the results from the survey in October 1993 to approximate infestation at *t*-1 for the survey in November 1994.

The fate of an infested site (persistence or extinction of its bug population) between survey *t* and *t*+1 could also influence to what degree it contributed to new establishment events in that time interval. If dispersal was concentrated at the beginning of the intervals between the surveys (shortly after each survey *t*), most infested sites infested at *t* contributed to new establishment events regardless of whether or not they went extinct before survey *t*+1. Conversely, if dispersal happened mainly towards the end of the intervals (shortly before each survey *t*+1), only sites that did not go extinct before survey *t*+1 were likely to contribute to new establishment events. If establishment was constant throughout the time intervals between surveys, the contribution of an infested site to new infestations was proportional to the time it was infested during a time interval.

Since we had no a priori knowledge about the timing of establishment, we compared the fit produced by three alternative ways to determine source sites at *t*. The first two methods were to either include *all* sites that were infested at survey *t* (Table 1, alternative II a) or only sites that were infested at survey *t and t*+1 (Table 1, alternative II b), corresponding to establishment at the beginning or the end of the time interval, respectively.

The source sites for the second alternative (Table 1, alternative II b) could include some sites that went extinct and experienced new establishment between *t* and *t*+1. Increased establishment that was independent of the level of infestation could therefore increase the number of sites we classified as source sites and the number of establishment events thus inducing a positive correlation between the response and explanatory variables. A comparison of the fit produced by model alternative II b (Table 1) with other alternatives therefore cannot prove but can only disprove dispersal at the end of the time intervals.

The third alternative we considered was continuous dispersal (Table 1, alternative II c). To describe this process, we had to determine the time of extinction between the surveys *t* and *t*+1. We were able to estimate the time of extinction only for infested sites sprayed between *t* and *t*+1 since spraying dates were known and extinction was likely immediately to follow insecticide spraying. Extinction dates of unsprayed infested sites were unknown. We therefore multiplied each site that was infested at survey *t* and sprayed between survey *t* and *t*+1 by the ratio of time between survey *t* and the spraying date over the time interval between survey *t* and *t*+1. Unsprayed sites that were infested at survey *t* were multiplied by 0.5 if they were uninfested at *t*+1 and by 1 if they were infested at *t*+1.

In summary, for a site infested at time *t*, infestation at each of the two surveys *t*-1 and *t*+1 could influence the potential of the site to act as source of establishment between *t* and *t*+1 in the three alternative ways. The three ways of treating time lags, described above, combined with these three ways infestation could influence establishment, led to nine ways of determining source sites at *t*.

The methods so far determined which of the sites infested at *t* was a source site between *t* and *t*+1. Depending on whether the number of infested sites or the number of bugs from these sites was a better predictor of the number of dispersers, one could either sum these weights directly (Table 1, alternative III a) or sum the product of each weight with the number of bugs found at the respective site and survey (Table 1, alternative III b).

The number of bugs seemed the more natural predictor, but if the number of bugs found at a site was an unreliable indicator of population size at a site or if the fraction of the population dispersing decreased with population size, as shown by McEwen, Lehane & Whitaker (1993), the number of sites could predict better. We used both alternatives in combination with all nine ways of determining a source site, obtaining 18 ways of calculating the explanatory variable.

Finally we investigated whether sites belonging to different ecotypes acted differently as source sites. Similar to the methods of Ceballos et al. (2005), we grouped sites into corrals, chicken-associated sites and domestic sites. Even though other studies found evidence for differences between goat and pig corrals (Cecere et al. 2004; Ceballos et al. 2005), we omitted this distinction to reduce the number of parameters to be estimated. We then calculated the explanatory variable either for all sites together (Table 1, alternative IV a), or for chicken associated sites and corrals separately (Table 1, alternative IV b).

Categorical variables

For each regression model we explored the effect of the categorical variables "season," "village location" and "ecotype of target sites." A hypothesis-testing approach to multiple categorical variables in regression poses a problem of multiple comparisons (Burnham & Anderson 2002, page 129ff). We therefore used the AIC_c to compare the fit produced by a single regression line with the fit obtained by estimating a separate regression line per factor level of the categorical variable. We did this comparison for all combinations of factor levels of all categorical variables. Since we had three dichotomous categorical variables, we had eight (2^3) alternative model structures to compare.

The variables "season" and "village location" were structured as follows. There were ten observations for the intervals from November to May and the eight from May to November. The two village locations (Amamá and Mercedes-Trinidad) each had all of these seasonal observations.

An effect of both "season" and "village location" could be expected a priori. Bug activity strongly fluctuates seasonally in the study region: bugs are much less active in the winter (Cohen & Gürtler 2001; Gorla & Schofield 1989). The two village locations differ in vegetation density. The third categorical variable was created by distinguishing three types of target sites (including domestic sites as a distinct type), thereby tripling the number of observations when we distinguished among ecotypes.

As mentioned above, for domestic sites only sensor box data (rather than flushing out data) were collected at half-year intervals. We included sensor box data from domestic sites as an additional response variable when we tested for ecotype differences in target sites but otherwise used the analysis of flushing out observations from peridomiciliary sites only. Again we compared the AIC_c of the model with three regression lines against AIC_c of the single-line-model and combined this comparison with all the comparisons mentioned above.

The number of parameters became too large when three factor level distinctions were combined with two explanatory variables. We therefore combined models that fitted two explanatory variables only with each categorical variable by itself or the combination of season or village location with target site ecotype since the distinction among target site ecotypes is nested within all other covariates. We therefore compared 504 alternative models altogether (all combinations of two response variables, 18 single explanatory variables and eight factor level combinations plus all combinations of two response variables, 18 double explanatory variables and six factor level combinations).

References:

Burnham, K.P. & Anderson, D.R. (2002) *Model selection and multimodel inference. A practical information theoretic approach.* Springer-Verlag, New York.

Ceballos, L.A., Vazquez-Prokopec, G.M., Cecere, M.C., Marcet, P.L. & Gürtler, R.E. (2005) Feeding rates, nutritional status and flight dispersal potential of peridomestic populations of *Triatoma infestans* in rural northwestern Argentina. *Acta Tropica*, **95**, 149-159.

Cecere, M.C., Canale, D.M. & Gürtler R.E. (2003) Effects of refuge availability on the population dynamics of *Triatoma infestans* in central Argentina. *Journal of Applied Ecology*, **40**, 742-756.

Cecere, M.C., Vazquez-Prokopec, D.G.M., Gürtler, R.E. & Kitron, U. (2004) Spatio-temporal analysis of reinfestation by *Triatoma infestans* (Hemiptera: Reduviidae) following insecticide spraying in a rural community in Northwestern Argentina. *American Journal of Tropical Medicine and Hygiene*, **71**, 803-810.

Cohen, J.E. & Gürtler, R.E. (2001) Modeling household transmission of American Trypanosomiasis. *Science*, **293**, 694-698.

Gorla, D.E. & Schofield, C.J. (1989) Population dynamics of *Triatoma infestans* under natural climatic conditions in the Argentine Chaco. *Medical and Veterinary Entomology*, **3**, 179-194.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman and Hall, London.

McEwen, P.K., Lehane, M.J. & Whitaker, C.J. (1993) The effect of adult population density on flight initiation in *Triatoma infestans* (Klug) (Hem. Reduviidae). *Journal of Applied Entomology*, **116**, 321-325.

Pitman, J. (1997) *Probability*. Springer Verlag, New York.

Rabinovich, J.E. (1972) Vital statistics of Triatominae (Hemiptera: Reduviidae) under laboratory conditions I. *Triatoma infestans* Klug. *Journal of Medical Entomology*, **9**, 331-370.