Crime Science

Open Access

CrossMark

A simple weighted displacement difference test to evaluate place based crime interventions

Andrew P. Wheeler[1*] and Jerry H. Ratcliffe[2]

Abstract

This work proposes a new statistic, the *weighted displacement difference* (WDD), as a simple test to evaluate crime reductions subsequent to place based interventions taking into account trends in comparison areas and potential spatial displacement or diffusion of benefits. The motivation is to have both a simple statistic that can give an estimate of the effectiveness of a crime reducing intervention, while also providing a measure of the standard error one would expect when examining crime counts simply by chance. The proposed statistic can be calculated in a spreadsheet with counts of crime before and after the intervention in both treated and displacement areas given matched control locations for each.

Code to replicate the analysis as well as an Excel® spreadsheet to calculate the WDD and its standard error can be downloaded from https://www.dropbox.com/s/rg7x9bg1o3d8nsf/WDD_Analysis.zip?dl=0

Keywords: Hot-spots, Policing, Intervention-analysis, Displacement, Diffusion-of-benefits

Background

The weighted displacement quotient (WDQ) is a simple statistic to identify whether a place-based intervention reduced crime in a treatment area relative to a control area, while taking into account potential spatial displacement of crime (Bowers and Johnson 2003; Guerette 2009; Ratcliffe n.d.). It has been used across a range of crime prevention evaluations, such as; burglary prevention (Bowers and Johnson 2003), estimating the impact of facilities such as public housing (Cahill 2011) or casinos (Johnson and Ratcliffe 2017) on crime rates, the crime prevention value of community support officers (Ariel et al. 2016), the crime reduction effects of closing brothels (Soto and Summers 2018), and whether there were any impacts on localized property and violence crime levels after the introduction of CCTV cameras (Ratcliffe et al. 2009).

Here we propose a related, but slightly different statistic, the *weighted displacement difference* (WDD). The motivation is to contribute a simple test statistic that provides an estimate of the uncertainty in the change of crime due to an intervention, under the assumption that the observed crime counts are Poisson distributed random variables. We aim to respond to the assessment of the WDQ's original authors (Bowers and Johnson 2003: 292) that "It would be useful to extend the analysis to include a test that enabled the researcher to identify significant levels of displacement". Although Bowers et al. (2011a) have written briefly about using a simulation approach to accomplish this with the original WDQ, both the new WDD test statistic we propose here, and its standard error, can be calculated in a spreadsheet. This statistic thus yields a simple tool for crime analysts to not only answer whether an intervention decreased crime, but to also say whether that decrease was larger than one would expect, given that crime counts will always fluctuate up or down by some amount simply by chance.

*Correspondence: apwheele@gmail.com
[1] School of Economic, Political, and Policy Sciences–Program in Criminology and Criminal Justice, The University of Texas at Dallas, 800 West Campbell Road, Mail Station GR 31, Richardson, TX 75080-3021, USA
Full list of author information is available at the end of the article

Springer Open

Table 1 Crime counts used for weighted displacement calculation

Area	Pre	Post
Treated	T_0	T_1
Control treated	Ct_0	Ct_1
Displacement	D_0	D_1
Control displacement	Cd_0	Cd_1

Methods

Table 1 explains the information needed to calculate the WDD. There are four areas; the treated area T, a control area comparable to the treated area Ct, a displacement area D, and a second control area comparable to the displacement area Cd. These are then indexed by counts of crime before the intervention took place with the subscript 0, and counts of crime in the post time period after the intervention took place with a subscript of 1. Note this is one more piece of additional information than the WDQ requires, a control area to correspond to the displacement area, which is sometimes considered in the evaluation literature but not frequently (Bowers et al. 2011a; Telep et al. 2014). We more critically discuss this unique requirement in a later section.

Based on this table, one can then identify the change in the treated, control, and displacement areas, subtracting the pre time period from the post time period.

$$
\begin{aligned}
\Delta T &= T_1 - T_0 \\
\Delta Ct &= Ct_1 - Ct_0 \\
\Delta D &= D_1 - D_0 \\
\Delta Cd &= Cd_1 - Cd_0
\end{aligned}
\tag{1}
$$

The WDD is then calculated as:

$$
WDD = \Delta T - \Delta Ct + \Delta D - \Delta Cd \tag{2}
$$

For a simple example, say crime decreased in the treated area by 10 (from 30 to 20), decreased in the control-treatment area by 5 (from 30 to 25), increased in the displacement area by 2 (from 10 to 12), and increased in the control-displacement area by 1 (from 11 to 12). In this example, the WDD would equal $-10 - (-5) + 2 - 1 = -4$. So although the treated area decreased by a larger margin, the control-treatment area also decreased (what might happen if crime is decreasing in the city overall). Because the displacement area also increased in crimes relative to the control-displacement area, the treatment effect is further reduced, but still has an estimate of decreasing 4 crimes in the combined treatment and displacement areas relative to their respective control trends.

Note in the case of spatial diffusion of benefits (Bowers et al. 2011b; Clarke and Weisburd 1994; Guerette and Bowers 2009), in which crime decreases were also observed in the displacement area relative to the control-displacement area, those additional crime reduction benefits would be added to the total crime reduction effects in the WDD. Thus the WDD is a not a direct analogy to the WDQ, but is more similar to the total net effects (TNE) estimate, which includes crime reductions (or increases) in both the treated and displacement areas in its calculation (Guerette 2009).[1]

Given these are counts of crime, assuming these counts follow a Poisson distribution is not unreasonable. While no set of real data will exactly follow a theoretical distribution, several prior applications have shown this Poisson assumption to not be unreasonable in practice (Maltz 1996; Wheeler 2016). We later discuss how departures from this Poisson assumption, in particular how crime tends to be over-dispersed compared to the Poisson distribution (Berk and MacDonald 2008), will subsequently affect the proposed test statistic.

The Poisson distribution has the simplifying assumption that its mean equals its variance, and hence even with only one observed count, one can estimate the variance of the WDD statistic (Ractliffe 1964). Making the assumption that each of the observations are independent,[2] one then can calculate the variance of the WDD as below, where $\mathbb{V}(X)$ is notation to be read as *the variance of X*[3]:

$$
V(\text{WDD}) = T_1 + T_0 + Ct_1 + Ct_0 + D_1 + D_0 + Cd_1 + Cd_0 \tag{3}
$$

Subsequently the test statistic to determine whether the observed WDD is likely due to chance given the null hypothesis that the WDD equals zero is then:

$$
\frac{\text{WDD}}{\sqrt{\mathbb{V}(\text{WDD})}} \sim \mathcal{N}(0, 1) \tag{4}
$$

---

[1] Using the same notation, $\text{WDQ} = (D_1/Ct_1 - D_0/Ct_0)/(T_1/Ct_1 - T_0/Ct_0)$, and $\text{TNE} = \left[T_0(Ct_1/Ct_0) - T_1 \right] + \left[D_0(Ct_1/Ct_0) - D_1 \right]$ (Guerette 2009: p. 41).

[2] To assume the treated and displacement areas are independent may seem on its face false. However, this assumption is equivalent to saying that the pre and post statistics for each area are drawn from the same underlying mean over time—which is what would occur if the intervention had no effect on crime. Hence this is a reasonable assumption on which to construct the null distribution of the test statistic.

[3] This is due to the fact that $\mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2 \cdot \mathbb{C}\text{ov}(X,Y)$, where $\mathbb{C}\text{ov}$ represents the *covariance* between X and Y. Assuming independence between the measures implies that the covariance term equals zero. Subsequently the standard error of the larger difference equation is simply the sum of the variances of the original Poisson counts. In the case this assumption is not met, one would likely expect the covariances between the terms to be positive, in which case the estimate of the variance *is too large*, and subsequently the test statistic will be conservative and *fail to reject* the null too frequently.

Table 2 Hypothetical crime control intervention and analysis

Area	Pre	Post
Treated	207	110
Control treated	308	318
Displacement	178	157
Control displacement	150	140

This test statistic, what we will refer to as Z_{WDD}, follows a standard normal distribution, e.g. ± 2 is a change that would only happen fewer than 5 times out of 100 tests if the null hypothesis of WDD equaling zero were true. One can then identify statistically significant decreases (or increases) in crime due to the intervention, while taking into account changes in the control areas as well as spatially displaced crime.

Note that the test statistic can still be applied even if one does not have a set of displacement and control-displacement areas. The user can simply set the displaced areas pre and post statistics to zero and conduct the test as is, since the zero counts do not add any extra variance into the statistic. This property can also be used to test the crime reduction effects in the treated and displaced areas independently (e.g. a crime reduction estimate in the treated area and a separate crime reduction estimate in the displaced area) if one so wishes.

The WDD test statistic is used, as opposed to the WDQ, as the ratios to calculate the quotient make the null distribution of the WDQ ill-defined, as the denominators can equal zero in some circumstances (Wheeler 2016). This property, especially for low crime counts, makes interpreting percentages very volatile (Wheeler and Kovandzic 2018). Examining the differences avoids this, but remains simple to understand and only requires slightly more work (identifying a matched control area for the displacement area) than the WDQ.

An example with discussion

Table 2 illustrates an example of calculating the WDD using an example adapted from the appendix of Guerette (2009)—examining a strategy aimed at reducing crime in rental units in Lancashire, UK. Because this example did not have a control-displacement area, we add fictitious numbers for illustration purposes.

Based on Table 2, the individual differences over time are:

$$\Delta T = 110 - 207 = -97$$
$$\Delta Ct = 318 - 308 = 10$$
$$\Delta D = 157 - 178 = -21$$
$$\Delta Cd = 140 - 150 = -10$$

And subsequently the overall WDD for this example is then:

$$\text{WDD} = -97 - 10 + (-21) - (-10) = -118$$

The variance of the WDD statistic then equals:

$$\mathbb{V}(\text{WDD}) = 207 + 110 + 308 + 318 + 178$$
$$+ 157 + 150 + 140 = 1568$$

And so the Z_{WDD} test statistic then equals:

$$Z_{WDD} = -118/\sqrt{1568} \approx -2.98$$

So in this case the reduction of a total of -118 crimes results in a statistically significant reduction. A Z-statistic of -2.98 has a two-tailed p value less than 0.003. The standard error of the test statistic (the square root of the variance), is approximately 40. One can use the same estimate of the standard error to estimate a confidence interval around the WDD statistic. In this case, a 95% confidence interval equals $-118 \pm 1.96 \cdot \sqrt{1568} \approx [-196, -40]$.

Simulation of the null distribution of Z_{WDD}

To ensure the validity of the proposed test statistic, simulations of the null distribution were conducted to establish whether the test statistic follows a standard normal distribution (i.e. has mean of zero and a variance of one). This was accomplished by simulating random deviates from a Poisson distribution of a given mean, and calculating the changes in each of the treatment, control-treated, displacement, and control-displacement areas. For each simulation the different areas were given the same underlying Poisson mean both in the pre and post treatment period, so there are no actual changes from pre to post in any of the areas. Any inferences from the simulation that concluded the intervention decreased (or increased) crimes would therefore be false positives. This procedure establishes that the false positive rate for the Z_{WDD} statistic is near where it should be if it were distributed according to the standard normal distribution. Values were simulated from Poisson means of 5, 15, 25, 50, and 100.

Additionally, we conducted a simulation run in which the baseline Poisson mean was allowed to vary between the treated and displacement areas, but was again constrained to be equal in the pre and post time periods. For this simulation run, the Poisson means were randomly drawn from a uniform distribution ranging from 5 to 100, so either treated or displacement areas could have larger baseline counts of crime. The simulation was then conducted for 1,000,000 runs, where one run would provide a single estimate of the WDD and the Z_{WDD} test statistic.

Table 3 Simulation results for no changes in underlying distribution

Poisson mean	p < 0.001	p < 0.01	p < 0.05	p > 0.95	p > 0.99	p > 0.999
5	0.0008	0.0101	0.0496	0.0492	0.0098	0.0008
15	0.0009	0.0100	0.0499	0.0493	0.0096	0.0009
25	0.0010	0.0100	0.0497	0.0504	0.0100	0.0009
50	0.0010	0.0098	0.0499	0.0501	0.0099	0.0010
100	0.0010	0.0100	0.0502	0.0500	0.0099	0.0010
Vary [5–100]	0.0010	0.0099	0.0503	0.0501	0.0099	0.0009
Z thresholds	Z < − 3.09	Z < − 2.33	Z < 1.64	Z > 1.64	Z > 2.33	Z > 3.09

Table 3 provides the test statistics for each of the extreme percentiles corresponding to the standard normal distribution. One can see that the WDD test statistic in the simulation near perfectly follows the standard normal distribution, and still provides a good fit even at the extreme quantiles. For example, one would expect a Z score of less than − 2.33 only 1% of the time according to the standard normal distribution with a mean of zero and a variance of one. In the simulation values this occurred at most 1.01% of the time (in the Poisson mean of 5), to a low of 0.98% of the time (with a Poisson mean of 15). One can read the other columns in a similar fashion and see that they are all quite close to the expected proportion of observations that should fall within each bin. Slight deviations are to be expected due to the stochastic nature of the simulation, but these statistics provide clear evidence that the null distribution of the test statistic does indeed follow a standard normal distribution. This is true both for the simulations in which the baseline mean number of crimes were constrained to be equal for both the treated and control areas, as well as the simulation run in which they were allowed to vary.[4]

One may ask why not simply use the current WDQ or TNE and attempt to place standard errors around that estimate? Unfortunately, this is not possible because the variance of the WDQ and TNE cannot be analytically defined in the same manner as the WDD. Both the WDQ and the TNE will have differing distributions under the null of no change depending on the baseline counts of crime. Additionally the extreme quantiles of the null distribution of no change for each of those statistics are quite volatile, mainly because the denominators in either of the statistics (in multiple places) can have very small values, causing the entire statistic to result in extremely large values (if defined at all) in relatively regular situations.[5] This point is further illustrated in Additional file 1: Appendix, where we calculate the WDQ and TNE statistics for the same simulated values as the WDD and show their volatility when no changes are occurring in the underlying means for either treated, control, or displacement areas.

While this simulation is only of a limited set of values, and in reality the treated and displacement areas are unlikely to have the same baseline Poisson mean, this is inconsequential for the WDD statistic (as is shown for the last simulation condition where they are allowed to vary). As long as the underlying crime counts follow a Poisson distribution, Z_{WDD} will closely follow a standard normal distribution. It is not necessary to assume that all of the treated and displacement areas have the same underlying mean for this to be true for the WDD, nor is it necessary to have differing suggestions for the strength of evidence given different baseline crime counts. This is not the case though with the WDQ nor the TNE. One cannot give generalizable advice (such as, a TNE statistic of + 50 is strong evidence that the intervention was effective in reducing crimes) as it depends on the baseline crime counts as to whether a swing of that size would be expected by chance. The same issue affects the WDQ statistic. It is however possible to give general advice when calculating the WDD and its associated Z_{WDD} value.

A note on the power of the WDD test

Due to the volatility of crime statistics, one should be wary that even if there is some evidence of a crime reduction, this test will often fail to reject the null hypothesis of no change over time. This is because it is necessary to

---

[4] Additionally, the mean and variance of each of the simulated distributions round to 0.00 and 1.00 respectively.

[5] Ratcliffe and Breen (2011) propose using the *Phi* coefficient and Soto and Summers (2018) use a Chi square and Cramer's V tests to see if the changes in crime counts are statistically different over time. Note this is different than asking for an estimate of the standard error around the WDQ statistic. So although a Phi coefficient, Chi-Square, or Cramer's V can potentially tell if there were changes over time, it cannot be used to infer either *directions* of the change, nor the *magnitude* of the change. The WDD statistic we propose is intended to accomplish both of those tasks.

Table 4 Recommendations for assessing evidence of crime reduction

Z_{WDD}	Strength of evidence	One tailed p-value
− 1.3	Weak evidence of reduction	0.1
− 1.6	Evidence of reduction	0.05
− 2.3	Strong evidence of reduction	0.01
− 3.1	Very strong evidence of reduction	0.001

have an intervention with effects larger than the inherent changes one would expect with count crime data, and in the case of lower baseline counts this can be a difficult threshold to meet. As an example of this type of *floor effect* (McCleary and Musheno 1981) consider this example: say the intervention, control-treated, displacement, and displacement-treated area each have around 5 crimes on average. This would result in a standard error estimate of the test statistic of $\sqrt{8 \cdot 5} \approx 6.3$. At a two-tailed alpha level of 0.05, one would need to have reduced *over 12* crimes (two times the standard error) to reject the null that the WDD was equal to zero. This is a near-impossible standard to meet with a baseline of only 5 crimes in the treated area. In that case one would need some combination of eliminating nearly all of the crimes in the treated and displacement areas and/or increases of crimes in the control areas to actually conclude that a statistically significant crime reduction took place. In the case of low base rates of crime, it will be rare to find strong evidence of crime reductions using the WDD statistic.

Table 4 displays a set of recommendations, relating one-tailed p-values to general strength of evidence concerning a crime reduction. Z_{WDD} values under − 1.3 (a one-tailed p-value of 0.1) provide weak evidence of crime reduction benefits to the intervention, with stronger evidence as one obtains lower values of Z. Z values over − 1.3 but below zero are additionally evidence of crime reductions, but could occur frequently by chance even if the intervention had no effect on crime. Note that these are intended to be general guidelines to crime analysts and police practitioners, many of whom have little formal training in statistics. While one could argue about the labels or the thresholds used to assign different levels of evidence, the overall goal is to take into account that a small negative WDD can simply occur by chance, and so one needs to assess both the size of that reduction as well as the expected error.

Aggregating up more crimes makes the potential to identify a statistically significant crime reduction more realistic. Figure 1 displays a set of curves that show the necessary crime reductions needed given a particular average number of crimes within each of the treated, displacement, and matched control areas. For example, at

an average of 100 crimes in each area, one needs to have a WDD of around 30 crimes to show weak evidence of effectiveness (those 30 crime reductions can count as a combination of the treatment area as well as diffusion of benefits in the displacement area). To obtain a one-tailed p-value of less than 0.05 it would take 40 or more crimes reduced. To reach the strong and very strong thresholds takes 57 and 76 crimes respectively. This is again a difficult standard even with a baseline of 100 crimes, as many interventions cannot be reasonably expected to reduce 100% of targeted crimes, and can only hope to reduce a sizable fraction in the best case scenario.

Because of this low power, one should keep in mind that failing to reject the null is not necessarily dispositive that the intervention was a failure. It could be the intervention is effective, it just does not prevent enough crimes to detect in the limited pre/post sample. Collecting more data over time, or expanding the intervention to additional areas, one might be able to identify reductions that appear to be larger than one would expect by chance.
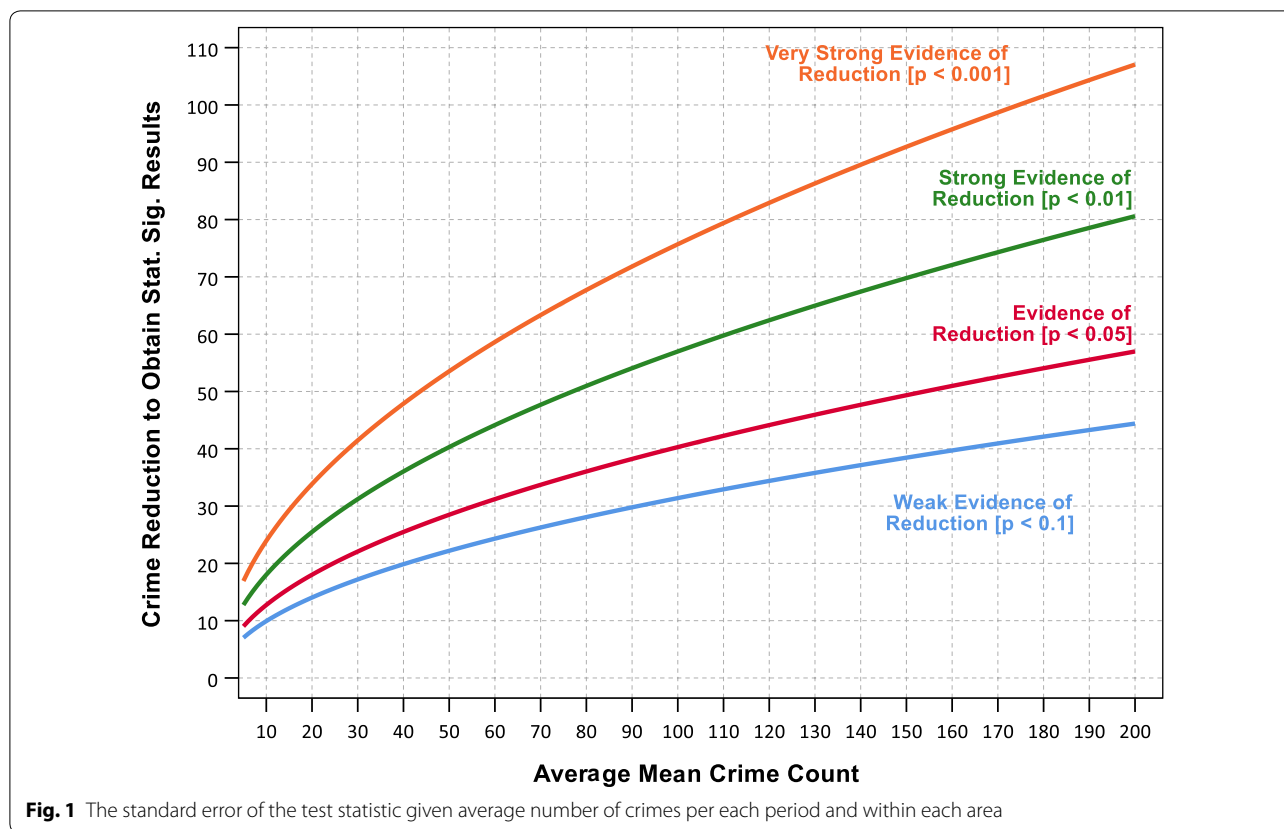
Choosing reasonable control and displacement areas

One important distinction between the WDD and the WDQ is that the control areas for both the treated and the displacement areas should be comparable in terms of the linear differences in crime counts over time. As such, it is important for the control areas to have similar counts of crime to the treated and displacement areas, a requirement that is not as strict for the WDQ.

For a simple example of where a control area would not appear to be reasonable is if crime decreased in the treated area from 100 to 90, and crime decreased in the control area from 10 to 9. Ignoring the displacement areas, although each area decreased a comparable *proportion* of crimes, this would result in a WDD statistic of − 9.

A general way to understand this assumption is that if the treatment area decreased by *n* crimes (without any intervention), would it be reasonable for the control area to also decrease by *n* crimes? With the control area having so few crimes to begin with, this assumption (going from 10 to 0) seems unlikely. But if one were starting from a baseline of 60 crimes, going from 60 to 50 is more plausible.

This would suggest generally that reasonable control areas will have similar *counts* of crime compared to the treatment area. Preferably control areas would also be close to equal in geographic size as the treated area, although if the treatment is focused on a hot spot of crime for smaller jurisdictions that only have one hot spot this may not be feasible. But this is not an absolute requirement though—the control areas merely need to follow the same linear trends over time as the

Fig. 1 The standard error of the test statistic given average number of crimes per each period and within each area

treatment area, which is likely to occur even if a control area is not a hot spot of crime. For example several long term evaluations of crime at micro places have shown that places with lower overall counts of crime tend to follow the same trajectory of crime decreases over time (Curman et al. 2015; Wheeler et al. 2016). In those cases, choosing several places in lower level trajectory groups to act as a control area for a high crime trajectory group location would be reasonable.

This also suggests displacement areas should have similar counts of crime compared to the treated/control areas if possible. While this is not necessary for the test statistic to behave appropriately as illustrated in the prior simulations, if one includes displacement areas that have substantially larger crime counts than the treated/control areas one will needlessly increase the variance of the test statistic, making it more difficult to detect if statistically significant crime reduction effects do occur. Typically one would assume displacement into only similar adjacent areas (Hesseling 1994), but if one wants to examine displacement into wider areas (Telep et al. 2014) it may make more sense to independently test the local and displacement effects, as opposed to pooling the areas into one test statistic.

With only counts of crime in the pre and post treatment periods, the linear change assumption cannot be

verified; however, this assumption can be approximately verified (or refuted) if the analyst has access to more historical data before the treatment took place. For example, if crime in the treated area in the three years prior to the intervention grew from 80 to 90 to 100, and crimes in the control area grew from 40 to 50 to 60, this would be *excellent* evidence that the linear differences assumption is appropriate. Both series appear to be in-sync with one another, are subject to the same external forces that may influence overall crime trends, and that the linear changes appear to be in lockstep for each series. This behavior represents a realistic counterfactual condition that is often sought by evaluators, in that the control area is not only similar to the treatment area in a pre-intervention cross-sectional analysis, but also longitudinally. If the control area however grew from 15 to 30 to 60, the linear changes of 10 crimes appears to be inappropriate. They are both trending upwards, but not in equivalent magnitudes.

Similarly if the control area grew from 8 to 9 to 10, the series are both growing, but are parallel in their *ratios*, not on a linear scale. The linear assumption can also hold for control areas that have higher baseline counts of the crime than the treated area. A control area going from 150 to 160 to 170 would also be excellent evidence that the linear differences assumption is appropriate, even

though it has a larger baseline count of crimes than the treated area.

Real world data are unlikely to be so convenient in verifying or refuting this assumption. Say if one had a control area that grew from 135 to 140 to 160, would the linear differences assumption be appropriate? While we cannot say for sure, one can keep in mind potential biases when choosing control areas. If you are choosing a control area that may change less than the treated area, in times of crime trending downward the intervention will likely appear *more* effective at reducing crime than it is in reality. If choosing a control area that may change less in times of crime trending upward the intervention will likely appear *less* effective than it is in reality. The obverse is true if choosing control areas that may change more than the treated and displacement areas when crime is trending downwards (will be more effective) or trending upwards (will be less effective). In times of no obvious trend of crime up or down, choosing control areas that have larger swings on average will introduce more variance in the statistic, but will not be consistently biased in either direction.

In the case that a single control area is reasonable as an overall measure of the expected crime change over time for both the treated and displacement areas, one can multiply the pre and post crime counts in the single control area by two and proceed with the statistic as usual. The WDD in that case would be $\Delta T - 2 \cdot \Delta Ct + \Delta D$. This might be reasonable if the treatment and displacement areas effectively cover an entire jurisdiction, have similar overall counts of crime, and control areas need to be drawn from a single neighboring jurisdiction. For all of these statistics, we recommend full and transparent disclosure of all the original counts (in addition to the differences over time) so that readers are able to appreciate any nuances in the data picture for all areas under examination, and factor this understanding into their interpretation of the study.

Conclusion and limitations

This essay provides a simple metric that analysts may use to evaluate place-based crime reduction interventions, while also providing an estimate of the standard error of that statistic. Prior work has shown that simple normal sample approximations have worked well when assessing the changes in two Poisson means (Detre and White 1970; Ractliffe 1964), and the simulations in this work also provide evidence that the test closely approximates the hypothetical standard normal distribution. This allows one to estimate the statistical significance of the intervention, as well as calculate confidence intervals around the effect estimate.

The biggest limitation of the current proposed test is that the control area needs to have relatively similar counts of crime that follow the same trends as the treated area for the WDD to be reasonable. The original WDQ was constructed to be more flexible because many times the treated and control areas could be different in their overall crime. Bowers and Johnson (2003) explicitly created a metric that would take into account those relative changes. Creating that relative metric though, through its use of ratios as opposed to differences, makes it impossible to place an estimate of the standard error around the WDQ given only one sample of treated, control, and displacement areas. While the WDD metric requires a compromise because the control area needs to have a commensurate number of crimes compared to the treated area to make sense, this is a step most analysts should attempt to adhere to when determining control areas when possible. If it is not the case that the control area has similar counts of crime or is not subject to the same independent influences on crime outside the scope of the intervention, it is questionable whether that control area is a reasonable selection to begin with.

While the proposed WDD test statistic has this additional burden of a control area for the displacement site, we do not believe this to be too cumbersome. Researchers frequently use simple buffers or adjacent locations as displacement areas for the treated location (Bowers et al. 2011a; Hesseling 1994), and that idea can be equally applied to create a displacement area for the control location (assuming the control area is far enough away that such buffers do not overlap). Geocentric buffers are not unambiguously the most appropriate way to construct displacement areas, though they are commonly applied to hot spots policing interventions. It is relatively easy to construct example buffer areas that cannot be expected to capture displaced crime—such as a buffer containing a body of water. It will always take a set of knowledge about the crime process under study as well as the local geography to construct a reasonable displacement area (Brantingham and Brantingham 2003; Ratcliffe and Breen 2011). As such, the process is no different for control area selection than it is for treatment area choices.

As the main motivation for proposing the test is to provide a simple tool for analysts to assess the effectiveness of a crime prevention intervention, it is also important to note that the test can easily be decomposed between the local effects and the spatial displacement/diffusion effects. Spatial displacement or diffusion of benefits are not inevitable in place based crime interventions (Telep et al. 2014). So although preferable to assess spatial displacement it should not be a hurdle that ultimately prevents even the assessment of the local effects of the intervention. It is likely the case that analysts will want to

assess the evidence of the local and diffusion effects independently for any intervention as well, which can easily be accomplished with the proposed WDD statistic.

An additional limitation of the WDD test is the power to identify statistically significant crime reductions. Although the null distribution of the test behaves appropriately with counts as low as five on average for each of the areas given our simulations, it is essentially impossible to conclude the intervention was effective when starting with such small baselines. This is, however, a problem endemic to all micro place-based policing research (Hinkle et al. 2013). One solution to this problem is to aggregate crimes over longer periods of time, or aggregate multiple treated areas.

A third limitation is the presumption that crime series follow a Poisson distribution. Reported crime counts are often over-dispersed, and often more closely follow a negative binomial distribution (Berk and MacDonald 2008). This will result in underestimates of the variance for the WDD, and subsequently cause more extreme Z-scores and increase false positives (saying the intervention caused crime to increase or decrease when in reality it had no effect on crime). Larger crime counts tend to more closely follow the Poisson approximation though (Osgood 2000), and similar analyses of actual crime data have illustrated the Poisson approximation is close enough even with low count crime data to construct error estimates over time (Wheeler 2016).

While admitting the test is not perfect due to these limitations, we believe it is reasonable enough to use in practice, and is better than the alternative of analysts using personal judgement as to whether the intervention is successful, based on subjective assessments of the magnitude of the changes in crime. The perfect need not be the enemy of the good.

In future work it is likely the case that the proposed test can be extended in several ways. Currently the test ignores the use of rates, which could be conceptualized as time (e.g. having a longer pre period than post period), area (assessing changes in crime density per area as opposed to crime counts), or as rates per victimizable targets (e.g. the number of residential burglaries per number of residences in each area). Other example tests of the differences in Poisson means have been adapted to take into account rates (Detre and White 1970; Krishnamoorthy and Thomson 2004), and so it seems likely the simple test provided here can also be adapted to estimating changes in rates. Finally, it is possible that exact tests could be developed with more power to identify changes in Poisson (or negative-binomial) distributed counts (Krishnamoorthy and Thomson 2004). But it is likely the case that the suggested normal based approximations will always be simpler to implement in software (including in

spreadsheets), thus making them much more accessible to a broader array of analysts.

Additional file

Additional file 1. Appendix.

Author details
[1] School of Economic, Political, and Policy Sciences–Program in Criminology and Criminal Justice, The University of Texas at Dallas, 800 West Campbell Road, Mail Station GR 31, Richardson, TX 75080-3021, USA. [2] Department of Criminal Justice, Temple University, Philadelphia, USA.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Ariel, B., Weinborn, C., & Sherman, L. W. (2016). "Soft" policing at hot spots—do police community support officers work? A randomized controlled trial. *Journal of Experimental Criminology, 12*(3), 277–317.

Berk, R., & MacDonald, J. (2008). Overdispersion and Poisson regression. *Journal of Quantitative Criminology, 24*(3), 269–284.

Bowers, K. J., & Johnson, S. D. (2003). Measuring the geographical displacement and diffusion of benefit effects of crime prevention activity. *Journal of Quantitative Criminology, 19*(3), 275–301.

Bowers, K.J., Johnson, S.D., Guerette, R.T., Summers, L., & Poynton, S. (2011a). Spatial displacement and diffusion of benefits among geographically focused policing initiatives. *Campbell Systematic Reviews 7(3).* https://www.campbellcollaboration.org/media/k2/attachments/0088_CCJG_Bowers_displacement.pdf, https://doi.org/10.4073/csr.2011.3

Bowers, K. J., Johnson, S. D., Guerette, R. T., Summers, L., & Poynton, S. (2011b). Spatial displacement and diffusion of benefits among geographically focused policing initiatives: A meta-analytical review. *Journal of Experimental Criminology, 7*(4), 347–374.

Brantingham, P. J., & Brantingham, P. L. (2003). Anticipating the displacement of crime using the principles of environmental criminology. *Crime Prevention Studies, 16,* 119–148.

Cahill, M. (2011). Using the Weighted Displacement Quotient to explore crime displacement from public housing redevelopment sites. *Cityscape, 13*(3), 103–134.

Clarke, R. V., & Weisburd, D. (1994). Diffusion of crime control benefits: Observations on the reverse of displacement. *Crime Prevention Studies, 2,* 165–183.

Curman, A. S. N., Andresen, M. A., & Brantingham, P. J. (2015). Crime and Place: A longitudinal examination of street segment patterns in Vancouver. *BC. Journal of Quantitative Criminology, 31*(1), 127–147.

Detre, K., & White, C. (1970). Note: The comparison of two Poisson-distributed observations. *Biometrics, 26*(4), 851–854.

Guerette, R. T. (2009). *Analyzing crime displacement and diffusion (Tool Guide Number 10)*. Washington, D.C.: Center for Problem Oriented Policing.

Guerette, R. T., & Bowers, K. J. (2009). Assessing the extent of crime displacement and diffusion of benefits: A review of situational crime prevention evaluations. *Criminology, 47*(4), 1331–1368.

Hesseling, R. B. P. (1994). Displacement: A review of the empirical literature. *Crime Prevention Studies, 3,* 197–230.

Hinkle, J. C., Weisburd, D., Famega, C., & Ready, J. (2013). The problem is not just sample size: The consequences of low base rates in policing experiments in smaller cities. *Evaluation Review, 37*(3–4), 213–238.

Johnson, L. T., & Ratcliffe, J. H. (2017). A partial test of the impact of a casino on neighborhood crime. *Security Journal, 30*(2), 437–453.

Krishnamoorthy, K., & Thomson, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference, 119*(1), 23–35.

Maltz, M. D. (1996). From Poisson to the present: applying operations research to problems of crime and justice. *Journal of Quantitative Criminology, 12*(1), 3–61.

McCleary, R., & Musheno, M. C. (1981). Floor effects in the time series quasi-experiment: Tracing the impacts of social policies. *Political Methodology, 7*(3/4), 181–203.

Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology, 16*(1), 21–43.

Ratcliffe, J.H. (n.d.) *WDQ spreadsheet calculator*. http://www.jratcliffe.net/software/wdq-spreadsheet-calculator/.

Ractliffe, J. F. (1964). The significance of the difference between two Poisson variables: An experimental investigation. *Journal of the Royal Statistical Society, Series C, 13*(2), 84–86.

Ratcliffe, J. H., & Breen, C. (2011). Crime diffusion and displacement: Measuring the side effects of police operations. *The Professional Geographer, 63*(2), 230–243.

Ratcliffe, J. H., Taniguchi, T., & Taylor, R. B. (2009). The crime reduction effects of public CCTV cameras: A multi-method spatial approach. *Justice Quarterly, 26*(4), 746–770.

Soto, C., & Summers, L. (2018). Macro-brothels as crime attractors: Impact of closure, displacement, and diffusion of benefits. *European Journal of Criminology* Online First. http://journals.sagepub.com/doi/abs/10.1177/1477370818788011?journalCode=euca

Telep, C. W., Weisburd, D., Gill, C. E., Vitter, Z., & Teichman, D. (2014). Displacement of crime and diffusion of crime control benefits in large-scale geographic areas: A systematic review. *Journal of Experimental Criminology, 10*(4), 515–548.

Wheeler, A. P. (2016). Tables and graphs for monitoring temporal crime trends: Translation theory into practical crime analysis advice. *International Journal of Police Science & Management, 18*(3), 159–172.

Wheeler, A. P., & Kovandzic, T. V. (2018). Monitoring volatile homicide trends across U.S. cities. *Homicide Studies, 22*(2), 119–144.

Wheeler, A. P., Worden, R. E., & McLean, S. J. (2016). Replicating group-based trajectory models of crime at micro-places in Albany, NY. *Journal of Quantitative Criminology, 32*(4), 589–612.