These are my thoughts on this right now, but note that Callaway and Sant'Anna are working on it, too, in a more formal way. Their paper, with a proposed estimator that solves some of these problems, is on its way!

Dose-response decomposition (of the *estimator*):

Consider a two-way fixed effects dose-response DD regression. d_i is the dose received by unit i, and $D_t \equiv 1\{t \geq t^*\}$, so there is no variation in *when* units received their dose (t^*), but there is variation in how much they get:

$$y_{it} = \alpha_{i\cdot} + \alpha_{\cdot t} + \beta^{DR} R_{it} + \epsilon_{it}$$

where $R_{it} \equiv d_i \times D_t$. Note that this is the same as a cross-sectional regression of the pre/post change in y on the dose. Define: $\Delta y_i \equiv \bar{y}_i^{POST} - \bar{y}_i^{PRE}$. Using the result that OLS is a weighted average of comparisons between observations you get:

$$\hat{\beta}^{DR} \equiv \frac{\sum_i \Delta y_i \left(d_i - \bar{d}\right)}{\sum_i \sum_t \left(d_i - \bar{d}\right)^2} = \frac{\sum_i \sum_{j>i} \left(\Delta y_i - \Delta y_j\right)\left(d_i - d_j\right)}{\sum_i \sum_{j>i} \left(d_i - d_j\right)^2} \qquad (1)$$

Now grouping these *observation* pairs by *dose* pairs you get a Wald/IV type result that expresses the coefficient as an average of Wald-DDs that compare two values of the dose only. Denote the two doses in each pair by ℓ for low and h, and use n_j for the share of units with the j^{th} dose. (These are key in a population-level analysis because the set of $P(d_i = j)$ are the treatment assignment model.) Also use $n_{jk} = \frac{n_j}{n_j + n_k}$ to denote the relative share of dose j in the (j, k) pair, which matters for the variance of d_i in the pair. Then write the coefficient as a weighted average of pairwise comparisons (ie. 2x2 DDs) scaled by their dose differences (ie. "Wald-DDs" in the language of Fuzzy DD, de Chaisemartin and D'Haultfoeuille):

$$\hat{\beta}^{DR} = \frac{\sum_\ell \sum_{h>\ell} n_h n_\ell \left(\Delta y_{ht} - \Delta y_{jt}\right)\left(d_h - d_\ell\right)}{\sum_\ell \sum_{j>i} n_h n_\ell (d_h - d_\ell)^2}$$

$$= \frac{\sum_\ell \sum_{h>\ell} \overbrace{(n_\ell + n_h)^2 n_{\ell h}(1 - n_{\ell h})(d_h - d_\ell)^2}^{subsample\,\&\,variance\,weight} \overbrace{\frac{\Delta y_{ht} - \Delta y_{\ell t}}{d_h - d_\ell}}^{\hat{\beta}_{h\ell}^{WDD}}}{\sum_\ell \sum_{j>i} (n_\ell + n_h)^2 n_{\ell h}(1 - n_{\ell h})(d_h - d_\ell)^2} \qquad (2)$$

$\hat{\beta}_{h\ell}^{WDD}$ are the "Wald DD" terms, and $s_{\ell h}$ are the scaled subsample/variance weights in (2), which gives the no-timing dose/response decomposition:

$$\hat{\beta}^{DR} \equiv \sum_\ell \sum_{h>\ell} s_{\ell h} \hat{\beta}_{h\ell}^{WDD} \qquad (3)$$

So the weighting comes from how far apart the doses are, the *relative* size of the two groups in each pair of doses, and the total size of the subsample.

What parameter does this identify and under what assumption?

Take (1) and replace $\Delta y_{it} = \Delta Y_{it}^0 + \frac{1}{T-(t^*-1)} \sum_{t \geq t^*} \left[Y_{it}^{d(i)} - Y_{it}^0\right]$. (The average across t is not that interesting here.) The expectation of Δy_{it} then for units with $d = k$ is $E[\Delta Y_{it}^0 | d = k] + E\left[Y_{it}^k - Y_{it}^0 | d = k\right]$

$$\Delta Y_{it}^{0k} + ATT(k|k)$$

This notation isn't perfect. I adapted it from Callaway and Sant'Anna, but it is mean to embody two things. We see the ATT at dose k, which is the first argument. There are potential outcomes $Y_{it}^0, Y_{it}^1, \dots, Y_{it}^K$ defined for every unit indexed by ordered treatment intensities (doses). But we only see this for units who actually have $d = k$, which here is denoted by the second k. We could define $ATT(k+1|k) \equiv E[Y_{it}^{k+1} - Y_{it}^0 | d = k]$ as the total ATT of dose $k+1$ among those units who actually have intensity k. Note that this is not the same as the per-unit effect of that marginal unit, which is the building block causal effect in Angrist and Imbens (1995) average causal response stuff. The per-unit effect between two doses $h > \ell$ for group k is: $ATT(h, \ell|k) \equiv E\left[Y_{it}^h - Y_{it}^\ell \middle| d = k\right] = E\left[Y_{it}^h - Y_{it}^0 \middle| d = k\right] - E\left[Y_{it}^\ell - Y_{it}^0 \middle| d = k\right] = ATT(h|k) - ATT(\ell|k)$.

Right now I think the easiest way to write the estimand is to express the ATT part as a variance-weighted sum like the decomposition and the counterfactual trends part:

$$\hat{\beta}^{DR} = \sum_{\ell} \sum_{h>\ell} s_{\ell h}\{ATT(h|h) - ATT(\ell|\ell)\} + \frac{\overbrace{\sum_i \Delta Y_{it}^0 \left(d_i - \bar{d}\right)}^{parallel\ trends}}{\sum_i \left(d_i - \bar{d}\right)^2} \tag{4}$$

The second term is "parallel" trends and here requires untreated counterfactual trends to be uncorrelated with the dose. Callaway and Sant'Anna call this a randomization-type assumption, and it also matches Goldsmith-Pinkham, Sorkin, and Swift's result for Bartik instruments that the "share" has to be exogenous.

I currently understand the parameter this way: add and subtract $ATT(\ell|h)$ to each term in curly brackets.

$$\frac{ATT(h|h) - ATT(\ell|h) + ATT(\ell|h) - ATT(\ell|\ell)}{d_h - d_\ell} = \frac{ATT(h,\ell|h)}{d_h - d_\ell} + \frac{ATT(\ell|h) - ATT(\ell|\ell)}{d_h - d_\ell} \tag{5}$$

The first term is the slope of the line connecting two points on the high-dose group's ATT function. By the mean value theorem this equals some derivative of that function (if it is continuously differentiable) in the interval $[d_h, d_\ell]$. That is a meaningful causal effect parameter.

The second term measures the extent to which heterogeneity in the dose-response function make the low-dose group a bad estimate of the low dose counterfactual for the high-dose population.[1] It is the difference in the "height" of the dose-response function at ℓ across the sub-populations (treatment effect heterogeneity at a given dose). If the low-dose units have better outcomes at low doses than the high-dose units would have had (a kind of Roy selection) this will be negative and each Wald DD is biased. Note, though, that it is scaled by $d_h - d_\ell$. A given difference in the sub-group ATTs at dose ℓ matters less when the different

[1] Can do it this way, too: add and subtract both $ATT(\ell|h)$ and $ATT(h|\ell)$ to the term in curly brackets yields and average of the dose-response function slopes for both groups plus a different bias term that comes from different heterogeneity, but I don't understand it right now:

$$\overbrace{ATT(h|h) - ATT(\ell|h)}^{ATT(h,l|h)} + ATT(\ell|h) - ATT(h|\ell) + \overbrace{ATT(h|\ell) - ATT(\ell|\ell)}^{ATT(h,\ell|\ell)}$$

$$= \sum_{\ell} \sum_{h>\ell} s_{\ell h} \left\{ \frac{ATT(h,l|h)}{d_h - d_\ell} + \frac{ATT(h,\ell|\ell)}{d_h - d_\ell} + \frac{ATT(\ell|h) - ATT(h|\ell)}{d_h - d_\ell} \right\}$$

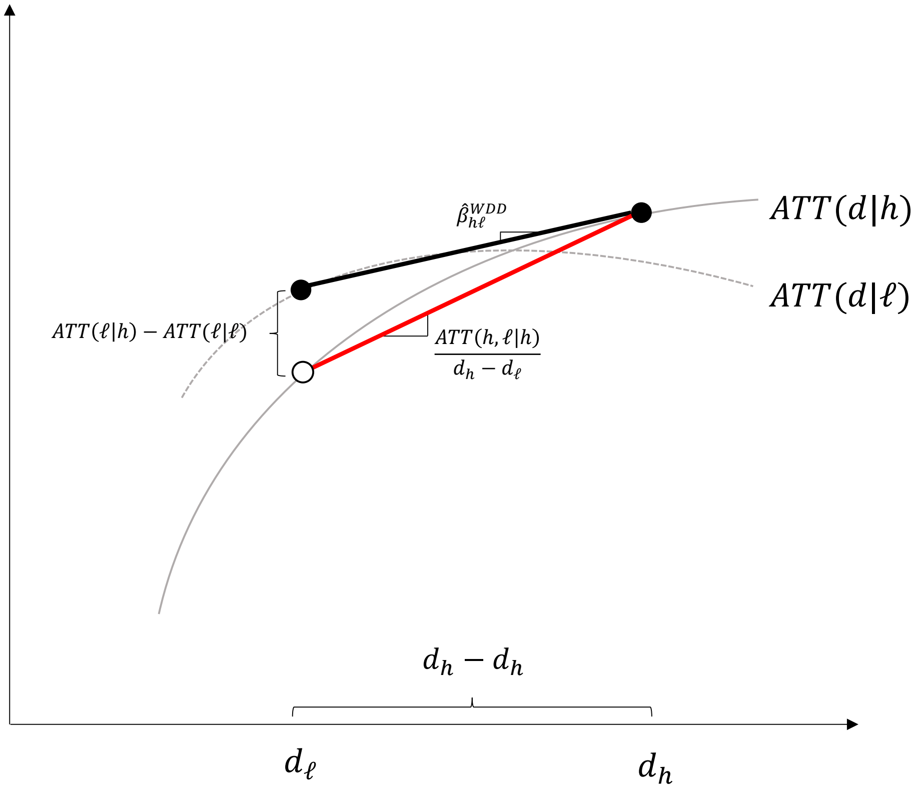
in doses is large. Of course, the two are connected because the largest differences in doses (denominator) may indicate the largest differences in the ATT functions.

This analysis for a given Wald DD is important because it refines the Fuzzy DD interpretation. There, we are worried because it is possible to compare two groups with different doses and get a wrong signed estimate. All ATT's could be positive, but the estimate may be negative. This shows that that's not necessarily wrong if the dose-response function is just non-monotonic. It could be true that there is an optimal dose beyond which the treatment is harmful. Then comparing a dose that is too high to a dose that is just right will yield a negative number that represents the marginal effects of those last $d_h - d_\ell$ dosage units.

But the flip side of that is: the source of heterogeneity matters for whether an estimate yields any meaningful causal effect. If there is a lot of heterogeneity in the dose-response function itself and it is correlated with the dose, then low-dose units' outcomes at low doses do not represent what high-dose units' outcomes would have been at those same dose levels.

Finally, there is again the weird OLS weighting. Common doses matter a lot: $(n_h + n_\ell)^2$. Comparisons between similarly common doses (ie. subsamples that aren't very skewed toward one dose) matter more: $n_{\ell h}(1 - n_{\ell h})$. And high-variance comparisons between very different dose units will matter most: $(d_h - d_\ell)^2$. Even if there were no heterogeneity in the dose-response function ($ATT(d; k) = ATT(d)$), this will give a parameter whose interpretation comes largely from the fact of having used OLS. Note, too, that this is just like average causal response stuff and likely is actually exactly the same in a sense.

This is the figure I've used to sort this out in my mind. It shows a single Wald-DD estimate and the underlying dose-response functions that give rise to it.


$\hat{\beta}_{h\ell}^{WDD}$
$ATT(d|h)$
$ATT(d|\ell)$
$ATT(\ell|h) - ATT(\ell|\ell)$
$\frac{ATT(h,\ell|h)}{d_h - d_\ell}$
$d_h - d_h$
d_ℓ
d_h