**METHODS AND MATERIALS**

***Literature Search***

We conducted a systematic search of relevant literature using ISI Web of Knowledge (WoS, available online). An initial search for literature was conducted in 2017 November using the following Boolean string: ("crop rotation\*" OR "crop sequence\*" OR "multiple cropping\*") AND (management\* OR control\*) AND ("weed biomass\*" OR "weed density\*" OR "weed seed bank\*") NOT (orchard\* OR vineyard\* OR agroforestry\*). This resulted in a total of 408 articles. An additional search was conducted in January 2018, using a slightly modified string, which included the word “organic” into the first phrase: ("crop rotation\*" OR "crop sequence\*" OR "multiple cropping\*" OR “organic”) AND (management\* OR control\*) AND ("weed biomass\*" OR "weed density\*" OR "weed seed bank\*") NOT (orchard\* OR vineyard\* OR agroforestry\*). This resulted in a total of 868 articles. An additional 10 records, coming from the authors collections of relevant literature, were screened for their eligibility. After removing duplicates, this resulted in a total of 885 articles, whose titles and abstracts were screened for eligibility. After this screening process, 102 full-text articles were screened for eligibility and **69** were selected. A full list of these publications is provided in S1 within the appendix. **Number is now different.**

No geographical or language restrictions were applied to the screening process, and the search period in ISI WoS, in both cases, was “All years” (i.e. 1864-2017/2018). Articles were screened using guidelines from Koricheva and Gurevitch (2014) based on the following criteria: i.) articles were published in peer-reviewed journals (as opposed to book chapters or conference abstracts) ii.) articles were conducted with structured experimental designs in a field setting iii.) articles reported one or both of the following response variables: weed biomass and/or weed density iv.) studies included, at least, one comparison of weed response from a simplified rotation (*SR*, control) and an diversified rotation (*DR*, treatment). We restricted our selection process to studies that used harvested crops (I.e. grain, fruit or forage biomass) and avoided studies in which the use of cover crops was stated. The literature search process detailed above can be found in the appendix (S2) in the form of a PRISMA flow diagram (Moher et al., 2009).

***Data extraction and processing***

We were interested in understanding the effect on weeds of adding crop species to monocrops (e.g. continuous maize) and repeating two-crop rotations (e.g. maize-soy sequence), as both these represent what we considered to be salient examples of SR within contemporary agricultural production. For the purposes of identifying and extracting relevant data, SR/DR comparisons were determined on a per-paper basis using crop species richness (the number of crops present over the course of a given rotation) as a guide. This resulted in the determination of the following SR types:

a.) SR = monocrop, species richness = 1 (e.g. continuous maize)

b.) SR = monocrop with alternating years of fallow, species richness = 1 (e.g. wheat grown in alternating years with a managed fallow)

c.) SR = a short two-year rotation (species richness = 2, *over two years*) (e.g. maize and soybean grown repeatedly in sequence)

If SR type a. or b. were present, all extracted comparisons used these as the control. In these instances, the DR was any crop rotation with a species richness ≥ 2. In **Y** papers, neither SR type a. or b. were present. Comparisons extracted from these papers used SR type c. as a control. In these cases, the DR was any crop rotation with species richness ≥ 3. These comparisons and their associated sample sizes are presented in table S3 within the appendix. ***Would a pie chart or graphic be better to give idea of relative make-up? I worry that because our sample sizes change so much via processing that these numbers become confusing if given in tables.***

Data were extracted from selected articles directly from tables, from figures using a digitization software (GetData graph digitizer, http:// http://getdata-graph-digitizer.com/), or from data sets provided by authors. Weed response variables were biomass (converted to g m-2, if not reported as such) or density (plants m-2). Data was taken only from studies in which weed response was measured within the same crop, to ensure legitimate comparisons. If a set of articles reported on weed response data for the same experiment documented over multiple years, we took care to extract data from the most recent publication. If a study provided weed response data for multiple time points *within* a given site-year, data from the last reported measurement were extracted. Data were also reported in three different ways with respect to measurements *between* or *among* site-years: at an end timepoint (*last*), averaged over multiple years (*average*), or represented by values from multiple years (*repeated measures*). To create a data set with as much uniformity as possible, we transformed *repeated measures* data points into *last* and *average* response data points. This was based on the length of the study in relation to the length of the longest rotation; if data were measured after the study had been in place more than two rotation cycles, the responses were averaged, and the number of replicates associated with the data were adjusted accordingly. For data measured in studies less than two rotation cycles old, the last year’s measurement was taken. The sample sizes before and after this processing step are detailed in S4 within the appendix. Additionally, in several studies, factors were analyzed that were not relevant to our hypotheses (e.g. planting date, herbicide rate). In these instances, data were pooled over levels of the factor with no increase in effective replication. **These processes resulted in a total of 60 observations for the response variable weed biomass, 98 for weed seedbank, and 237 for weed density. *This will change based on seed bank study elimination etc.***

For each data point, publication information, study location, experimental design, study age, and the number of replicates were recorded. Information related to crop and weed management was also extracted from the studies based on our hypotheses, and their availability. Much of this information was accessible directly from the papers themselves. However, specific to planting date information, if planting dates were explicitly not described, approximate planting months were obtained by contacting authors, grey literature and extension guidelines. A clear list of the moderators developed from this extracted information is enumerated in greater detail in S5 within the appendix.

***Data analysis***

To estimate effect sizes, we calculated the logarithm of the ratio of the weed response in the ER to the response in the SR (Eqn X) (Gurevitch et al., 2018). (Borenstein et al., 2009; Lajeunesse, 2014).

LRR = ln(XT/XC) Eqn X

Where XT = treatment (ER) mean and XC = control (SR) mean.

In 11 cases either the treatment or control value was zero – these points were removed from the analysis because adding an arbitrary amount to all values can result in comparisons changing from negative to positive, and can result in unrealistic effect sizes.

Only 9 of the studies reported variances in some form. As such, we used non-parametric weighting based on sample sizes (Lajeunesse, 2014):

σ2 (LRR) = nTnC/(nT + nC) Eqn X

where nT = treated (ER) group sample size and nC = control (SR) group sample size.

To estimate effect sizes, we fit a linear mixed-model with study as a random effect and weighting as described above using the lme4 package (cite it, Pinheiro I believe). A separate model was fit to each response variable. Initial analyses was conducted without moderators to assess overall effects. To evaluate the effect of apriori chosen moderators, mixed-effect models were fit for response using each moderator as a single fixed effect, study as random, and non-parametric weighting (Eqn X). Confidence intervals around estimated mean effects were also bootstrapped using 2500 iterations, maintaining sample sizes within each moderator level to ensure proportional representation was maintained (Adams et al. 1997).

*Adams, D. C., Gurevitch, J., & Rosenberg, M. S. (1997). Resampling tests for meta‐analysis of ecological data. Ecology, 78(4), 1277-1283.*

Mean LRR and CI were all back-transformed, and are expressed as a percentage change from the control (SR).

To ensure the robustness of our results, we tested the sensitivity of mean effect sizes to included studies by performing a leave-one-out analysis (Philibert et al., 2012; Gurevitch et al, 2018), using the mixed model to estimate means and confidence intervals for each subset of data.

**Other**

**S1 – pub list**

**S2 – PRISMA lit search**

**S3 – SR/DR table with sample sizes etc. OR figure of some sort**

**S4 – Data processing before/after sample sizes**

**S5 – Moderator table**