Article:

Evans, C.R., Borrell, L.N., Bell, A. orcid.org/0000-0002-8268-5853 et al. (3 more authors) (2024) Clarifications on the intersectional MAIHDA approach: a conceptual guide and response to Wilkes and Karimi (2024). Social Science and Medicine, 350. 116898. ISSN 0277-9536

https://doi.org/10.1016/j.socscimed.2024.116898

Clarifications on the Intersectional MAIHDA Approach:

A conceptual guide and response to Wilkes and Karimi (2024)

Authors:

Clare R. Evans (corresponding author)[1]

Luisa N. Borrell[2]

Andrew Bell[3]

Daniel Holman[4]

S.V. Subramanian[5,6]

George Leckie[7]

[1] Department of Sociology, University of Oregon, Eugene, OR, USA.

Email: cevans@uoregon.edu. ORCID: https://orcid.org/0000-0002-9862-9506

[2] Department of Epidemiology & Biostatistics, Graduate School of Public Health & Health Policy, The City University of New York, New York, NY, USA.

ORCID: https://orcid.org/0000-0002-0560-4853

[3] Sheffield Methods Institute, University of Sheffield, Sheffield, UK. ORCID: https://orcid.org/0000-0002-8268-5853

[4] Department of Sociological Studies, University of Sheffield, Sheffield, UK. ORCID: https://orcid.org/0000-0003-4134-6238

[5] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ORCID: https://orcid.org/0000-0003-2365-4165

[6] Harvard Center for Population and Development Studies, Cambridge, MA, USA.

[7] Centre for Multilevel Modelling and School of Education, University of Bristol, Bristol, UK.

ORCID: https://orcid.org/0000-0003-1486-745X

Abstract

Intersectional Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) has been welcomed as a new gold standard for quantitative evaluation of intersectional inequalities, and it is being rapidly adopted across the health and social sciences.

In their commentary "What does the MAIHDA method explain?", Wilkes and Karimi (2024) raise methodological concerns with this approach, leading them to advocate for the continued use of conventional single-level linear regression models with fixed-effects interaction parameters for quantitative intersectional analysis.

In this response, we systematically address these concerns, and ultimately find them to be unfounded, arising from a series of subtle but important misunderstandings of the MAIHDA approach and literature. Since readers new to MAIHDA may share confusion on these points, we take this opportunity to provide clarifications.

Our response is organized around four important clarifications: (1) At what level are the additive main effect variables defined in intersectional MAIHDA models? (2) Do MAIHDA models have problems with collinearity? (3) Why does the Variance Partitioning Coefficient (VPC) tend to be small, and the Proportion Change in Variance (PCV) tend to be large in MAIHDA? and (4) What are the goals of MAIHDA analysis?

Funding

This work was funded in part by the Economic and Social Research Council, grant number ES/X011313/1.

Highlights:

- MAIHDA is a new gold standard for estimating inequalities intersectionally.
- A recent critique argues MAIHDA has conceptual and collinearity issues.
- We examine these concerns and conclude they are unfounded.
- We provide clarifications about MAIHDA model specification and interpretation.
- We elucidate the fundamental goals of MAIHDA analyses.

Clarifications on the Intersectional MAIHDA Approach:

A conceptual guide and response to Wilkes and Karimi (2024)

1. Introduction

In their commentary "What does the MAIHDA method explain?", Wilkes and Karimi (Wilkes & Karimi, 2024) (henceforth WK) raise concerns about the quantitative approach known as intersectional Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA), a new application of multilevel models first proposed by Evans (2015) (later published by Evans et al. (2018)) and explored and expanded by others (Axelsson Fisk et al., 2018; Bell et al., 2019; Evans, 2019c; Jones et al., 2016). MAIHDA has been welcomed as a "new gold standard" for quantitative evaluation of inequalities (Merlo, 2018), and it is being rapidly adopted across the health and social sciences (Alvarez & Evans, 2021; Cubells et al., 2024; Evans, 2024; Keller et al., 2023; Moreno-Agostino et al., 2023). It has been validated by independent teams using both empirical (Bell et al., 2023; Evans, 2019a; Holman et al., 2020) and simulation-based analyses (Bell et al., 2019; Evans et al., 2018, 2020; Mahendran et al., 2022b, 2022a).

In their commentary, WK argue that MAIHDA suffers from methodological issues, especially collinearity in its core structure, and advocate for the continued use of conventional single-level linear regression models with fixed-effects interaction parameters for quantitative intersectional analysis. They claim their critique is "joining a growing chorus (Block et al., 2023; Lizotte et al., 2020; Yang, 2023) advocating for the ongoing use of single-level models for the quantitative study of intersectionality" (Wilkes & Karimi, 2024, p. 9). However, this makes it appear that critiques of MAIHDA are stacking up, which is not the case. Notably, of the three publications they cited, one did not address MAIHDA at all, and therefore cannot be said to be arguing against its use (Block et al., 2023). In the second (Lizotte et al., 2020), the authors raised some initial questions about whether the MAIHDA approach compares favorably with single-level models, and these questions have been satisfactorily addressed (Evans et al., 2020). In fact, the same authors subsequently conducted a systematic review (Bauer et al., 2021) and extensive simulation studies comparing MAIHDA with alternative modeling approaches (Mahendran et al., 2022b, 2022a), leading them to change their initial assessment, now concluding that MAIHDA “holds promise as a statistically efficient method for predicting outcomes across large numbers of intersections” (Bauer et al., 2021, p. 9) and “is the preferred unbiased method for accurate estimation of high-dimensional intersections at smaller sample sizes” (Mahendran et al., 2022b, p. 1).

The final cited publication (Yang, 2023) discussed similar questions about MAIHDA to those articulated by WK, namely, around the meaning of models that partition variance across additive (fixed) terms and interaction (residual random) terms. Because Yang (2023) only addressed MAIHDA briefly in the background section, explaining their reasons for not using it, we focus our response on the WK commentary, while addressing the points raised by Yang (2023) as well.

In this response, we systematically address WK’s concerns, and ultimately find them to be unfounded. We argue that these concerns arise from a series of subtle but important misunderstandings of the MAIHDA approach and literature. Since readers new to MAIHDA may share confusion regarding these points, we take this opportunity to clarify them.

After providing a brief review of the intersectional MAIHDA approach, we organize our response around four important clarifications: (1) At what level are the additive main effect variables defined in intersectional MAIHDA models? (2) Do MAIHDA models have collinearity problems? (3) Why does the Variance Partitioning Coefficient (VPC) tend to be small, and the Proportion Change in Variance (PCV) tend to be large in MAIHDA? and (4) What are the goals of MAIHDA analysis? For additional details on the approach, we recommend a recent tutorial on MAIHDA (Evans et al., 2024).

2. A Brief Review of the Intersectional MAIHDA Approach

Intersectional MAIHDA (an application of MAIHDA to the analysis of intersectional inequalities) refers to an analytic approach in which a two-level multilevel regression model is specified, with individuals (level-1) nested hierarchically in intersectional social strata (level-2). The broad purpose of MAIHDA is to quantify inequalities across diverse intersectional strata, in terms of both predicted (average) values and variation/heterogeneity within and between strata. In MAIHDA, these social strata typically correspond to the unique combinations of social identities, positionalities, or material conditions of interest, such as gender, race/ethnicity, and socioeconomic status. For instance, one stratum may be specific to high-income Black women, another to middle-income white men, and so on for all combinations of the relevant variables. The choice of categories and axes of marginalization to be included in an analysis should be guided by theory, prior empirical work, and practical limitations of the data.

Early work on MAIHDA discussed how variation in mean outcomes for individuals who share the same intersectional positionality is approximately analogous to how multilevel models conventionally allow us to model variation in mean outcomes among respondents who reside in the same neighborhood, attend the same school, or are employed in the same workplace. In this sense, a “social stratum” can be conceptualized as a clustering unit, with everyone in the stratum reporting identical combinations of the individual characteristics used to define the stratum (e.g., the same gender, race/ethnicity, and income band). Therefore, this subset of individual characteristics consists of level-2 (stratum-level) variables, rather than level-1 (individual-level) variables. These strata are theorized as (level-2) proxies for positionalities within a landscape of interlocking systems of oppression, marginalization, and inequality (e.g., sexism, racism, and socioeconomic inequality). Shared positionality may mean that respondents face similar forms of (dis)advantage on the basis of such positionality—thus creating potential inequalities between strata in their mean observed outcomes.

The two core MAIHDA models are the *null model* and the *additive main effects model*. For continuous outcomes, we generally recommend linear model specifications, and for binary or binomial count outcomes, logistic versions of MAIHDA analysis have been illustrated in the literature, which allow for the generation of stratum-specific predicted probabilities (Axelsson Fisk et al., 2018; Evans, 2019a; Evans et al., 2024; Mahendran et al., 2022b). Here, like WK, we focus on the linear case. In its linear specification, a null MAIHDA model is a two-level random-intercept linear regression model, where y_{ij} is the outcome for individual i ($i = 1, \ldots, n_j$) in social stratum j ($j = 1, .., J$). It can be written as:

$$y_{ij} = \beta_0 + u_j + e_{ij}$$

where $u_j \sim N(0, \sigma_u^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. In the null model β_0 is the precision-weighted mean of the stratum means and u_j is the stratum random effect for stratum *j*, which captures how different the mean in that stratum (given by $\beta_0 + u_j$) is from the overall mean β_0. Across all strata, the u_j are assumed to be normally distributed with mean of 0 and variance σ_u^2. The residual e_{ij} measures the deviation of the observed outcome for individual i in stratum j from the stratum mean. Similarly, e_{ij} is assumed to be normally distributed with mean 0 and variance σ_e^2.

The primary use of the null model is that it helps characterize the total inequality in mean outcomes across strata, with the stratum-level variance σ_u^2 capturing the between-stratum inequality. The practical importance of the magnitude of this between-stratum variance hinges partially on the amount of ‘background’ variation between individuals (potentially attributable to unmodelled characteristics) at level-1. The Variance Partition Coefficient (VPC) is calculated to evaluate the between-stratum variance standardized against such background variation. Specifically, the VPC is defined as the proportion of the total variance in y_{ij} (given by $\sigma_u^2 + \sigma_e^2$) that lies between-strata.

$$\text{VPC} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

The second model in MAIHDA is an additive main effects (ME) model, where we enter the set of categorical variables that define the level-2 strata (e.g., gender, race/ethnicity, and education level) as a series of level-2 fixed-effects dummy variables (i.e., the "main effects"):

$$y_{ij} = \beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj} + u_j + e_{ij}$$

where, again, $u_j \sim N(0, \sigma_u^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. The new terms $x_{1j}, \ldots, x_{pj}$ denote the p dummy variables and $\beta_1, \ldots, \beta_p$ are their associated regression coefficients. This model includes no fixed interaction parameters (e.g., we include dummy variables for 'woman' and 'Black' but no product term for 'woman × Black'). Instead, the net interaction effect between the variables for stratum j is captured by the stratum random effect u_j. Thus, the interpretation of u_j has changed from that in the null model, and now measures the deviation of the stratum means from the values expected by the additive main effects alone. The u_j continue to be assumed normally distributed with mean 0 and between-stratum variance σ_u^2. The latter measures the variance that remains between strata after adjustment for the additive main effects. In contrast, the residual e_{ij} has the same interpretation as in the null model, and continues to be assumed to be normally distributed with mean 0 and within-stratum-between-individual variance σ_e^2. Furthermore, the covariates are assumed independent of the stratum random effect.

The additive ME model has multiple applications. First, it can be used to produce total expected values (inclusive of fixed additive and random interactive effects) of the outcome in each stratum. The summation $\beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj}$ gives the expected outcome value for stratum j based on the additive main effects alone (more on interpretation of this later), while $\beta_0 + \beta_1 x_{1j} + \cdots + \beta_p x_{pj} + u_j$ gives the total expected outcome value for stratum j, based on additive and interaction parameters.

But why, if the goal is simply to calculate expected values of the outcome in all strata, do we not simply set up a high-dimensional cross-tabulation to identify all social strata of interest, and then calculate the (raw) average scores in each group? Conventional single-level approaches, including fully specified interaction models, accomplish this task using a regression model. A useful feature of the MAIHDA approach, and one that makes the added model complexity worthwhile, is that the predicted stratum means are “shrunken” predictions that pull the observed stratum mean values for smaller strata towards those predicted based on the additive main effects. This precision-weighting protects against the ‘small N problem’ where extreme mean values are calculated based on relatively few cases, leading to potential mischaracterizations of outcomes in strata. This is often an issue in this sort of analysis where we define a large number of strata, because some strata are multiply minoritized or composed of unusual intersections (e.g., lowest education level and highest income level), and so some strata may have few observations even in large datasets. Thus, MAIHDA improves the reliability and robustness of the predictions for these small strata (Bell et al., 2019; Evans et al., 2018; Mahendran et al., 2022b, 2022a) and helps to address the multiple testing problem (Bell et al., 2019). We can then examine the predicted u_j to identify strata whose predicted means are most different from that predicted by the additive main effects alone.

A second use of the additive ME model is that by partitioning the outcome’s inequality patterns across strata into additive (fixed effect) and interaction (random effect) parts, the ME model enables us to consider inequality patterns in terms of *universality-versus-specificity*. For instance, suppose we find a health disparity where Black respondents report worse health outcomes than white respondents (which we theorize as originating from social determinants, e.g., embodiment of experiences with race-based discrimination (Krieger, 2011; Phelan & Link, 2015)). Can we conclude that “the health disparity for Black versus white respondents is X” and have that conclusion stands for *all* Black and white respondents? In other words, is the inequality of size X “universal” across other axes of comparison? Or, alternatively, will the size of that inequality differ based on gender (e.g., is it larger for Black-women versus white-women than for Black-men versus white-men), or based on age, education, income, or disability? If we find that the inequality patterns observed are reasonably well explained by additive parameters, then the more general statement is an acceptable shorthand summary of the finding. If, however, there are meaningful deviations from the "general additive story," then we would find interaction effects to be significant for some (or all) strata, and we would have to be *specific* in our description of inequality patterns.

A useful statistic for describing the relative contribution of additive and interaction effects is the Proportional Change in Variance (PCV), calculated as:

$$\mathrm{PCV} = \frac{\sigma^2_{u(\text{Null Model})} - \sigma^2_{u(\text{Main Effects Model})}}{\sigma^2_{u(\text{Null Model})}}$$

The PCV measures how much the between-stratum variance changes between the null and ME models (e.g., how much of the between-stratum variance is explained by adding the main effects, with a higher percentage indicating that a greater share of the total between-stratum variance measured in the null model is accounted for by the additive main effects). In this way, the model allows us to uncover the extent to which intersectional inequalities are additive versus multiplicative.

3. Response to the Concerns Raised by Wilkes and Karimi (2024)

3.1. Clarification #1 – At what level are the additive main effects variables defined in intersectional MAIHDA models?

WK (2024) repeatedly and incorrectly described the ME model as one that included demographic variables, such as race, gender, and poverty status, as individual level-1 fixed effects. For instance, they define the model as one that "adds the individual level covariates (e.g., race, gender, poverty status) as additive main effects (βx_{ij}) at level 1" (p.2). Furthermore, they assert that the purpose of including these variables is "to consider the extent to which individual-level variables…explain the variation within and across intersectional strata (level 2)" (p.2).

However, as we discussed above, in the MAIDHA approach these variables operate as stratum-level (level-2) fixed effects, not individual-level (level-1) fixed effects. The MAIHDA literature has been clear on this point since the beginning: “Contrary to most multilevel models where social categories such as race and income are individual-level covariates, here such covariates are properties of the strata level” (Evans, 2015, p. 20; Evans et al., 2018, p. 67). For this reason, in the above equations and in published studies using MAIHDA, the ME dummy variables are indexed by j and not ij (e.g., x_{1j} not x_{1ij}, indicating a level-2 variable that is the same for all respondents in stratum j). WK appear to partially arrive at an understanding of this, as they later describe how all respondents in a stratum will share the same values of the ME variables (“everyone within each individual stratum is already only one race, one gender, and one poverty level” (Wilkes & Karimi, 2024, p. 9)). Thus, because they operate as level-2 variables, the ME dummy variables cannot be expected to explain within-stratum/between-individual (level-1) variance. However, because WK missed that these are level-2 variables, they interpreted this as a flaw with the method rather than an entirely expected property of a level-2 variable. Level-2 variables are not expected to account for level-1 variation when added to a multilevel model; they only explain level-2 variation.

This treatment may seem unusual to some readers, since variables such as age, sex/gender, and race/ethnicity are individual-specific characteristics. However, here we operate them as descriptors of groups (strata) and theorize them as proxies for the impact of common experiences with social forces (e.g., racism, not race). The term ‘strata’ here is intentionally used both to avoid the word group, which some scholars have critiqued as implying social cohesion (Brubaker, 2002), and to provide a conceptual link to stratified analyses, in which a researcher will impose stratification schemes for analytic purposes. While MAIHDA operates differently than a stratified analysis, the comparison is useful—strata are defined, and the characteristics used to define these analytic ‘groups’ cease to be individual-level variables, and become characteristics describing the group.

WK seem to take no issue with including additive and interaction parameters on the same level (level-1) in single-level models, and they even encourage the use of single-level models for this reason. Importantly, MAIHDA follows a conceptually similar approach—additive main effects (level-2) and stratum-level random effects (level-2) are specified at the same level, and decompose inequalities into additive and interactive components, albeit in a subtly different way than single-level models.

3.2. Clarification #2 – Do MAIHDA models have collinearity problems?

We now address the main question raised by WK about MAIHDA—does this approach have problems with collinearity? Collinearity is broadly understood to refer to a linear relationship existing between two or more explanatory (usually, fixed effect) variables in a regression model (Barrie Wetherill et al., 1986). WK describe strata in intersectional MAIHDA as a type of “composite measure”. They discuss how one would not wish to include, in a single-level model, both the composite measure and the variables used to construct it (e.g., separately including income, education, and a composite measure of income and education) as fixed effects, because this would result in exact collinearity. Thus far, we entirely agree with this logic. However, they then go on to describe how moving the stratum composite measure to level-2 as a random effect/clustering unit “in effect moves the potential collinearity between variables at one level to a potential collinearity between the second level and the variables at the first level. The potential collinearity is displaced rather than erased” (Wilkes & Karimi, 2024, p. 6). However, this is not the case.

First, and as we have explained in Section 3.1 (Clarification #1), the additive main effects they refer to are level-2 variables rather than level-1 variables. Consequently, a better analogy to how the MAIHDA main effects (ME) model operates is a two-level model of individuals (level-1) nested in neighborhoods (level-2), which also includes neighborhood characteristics as level-2 variables, such as a neighborhood-level measure of air pollution. In such a model, there is also "collinearity" between neighborhood characteristics and neighborhood identifiers. However, this collinearity is not a problem—in fact, it is a standard and expected feature of the multilevel modeling approach. Neighborhoods are clustering units (like strata), and there is variation between them in mean outcomes (again, like strata). Including a level-2 variable such as neighborhood pollution will "soak up" (or account for) some of this between-neighborhood variance. Similarly, in the MAIHDA additive ME model, including the level-2 variables that were used to construct the strata (e.g., woman, Black) will account for some of the between-stratum variance. We would not characterize the neighborhood example as having a collinearity problem, and neither should we in the MAIHDA case. In both models, we are simply using level-2 variable(s) to account for level-2 variance in the outcome, and partitioning the variance across the fixed and random parts of the model.

It also does not follow that modeling both the stratum-level and the variables that define the strata is "tautological" (Wilkes & Karimi, 2024, p. 3,5). Rather, it allows us to do something common in most statistical models: separating variance that has been explained by some variables from that which has not been explained.

3.3. Clarification #3 – Why does the VPC tend to be small, and the PCV tend to be large in MAIHDA?

WK erroneously state: "That an individual-level composite variable—strata—is specified as a level-2 context explains the lack of within group variance in MAIHDA" (Wilkes & Karimi, 2024, p. 7). Whilst it is true that the specification of the MAIHDA model will shift some variance (that would normally be considered "individual-level") into a second level, thereby reducing the level-1 individual-level variance, this statement also suggests that this level-1 variance is typically small in MAIHDA, and that is not the case.

Based on theory and past empirical work, we generally expect individual-level ("within group/stratum") variance to be large relative to the variance between strata because of the numerous explanations for differences (or variation) between individuals. At best, categorizations of populations such as "woman" or "Black" are rough proxies for highly complex and variable social experiences, with meanings that shift along other identity-axes and across time and context. Predictably, therefore, we see this reflected in the published literature using MAIHDA. Where the individual-level (level-1) variance is large relative to the stratum-level (level-2) variance, the VPC in the null model is expected to be small. In Table 1, we present results reported in recent studies using MAIHDA across a range of outcomes (including studies discussed by WK), and provide level-specific variances, the null model VPC, and the PCV whenever they were available. Consistent with similar findings from the literature on multilevel models of individuals clustered by neighborhoods (Subramanian & O'Malley, 2010), the VPC tends to be smaller than 10% and is frequently <5% (Evans, 2019a; Holman et al., 2020). In some instances, larger values are observed, but the majority of the variance is still typically found at the individual-level.

WK go on to describe how, in the examples they discuss, "there is *very little within stratum variance change* and *dramatic between stratum variance change"* (Wilkes & Karimi, 2024, p. 9) between the null and ME models. They view this large change in between stratum variance between the null and ME models with suspicion: "such dramatic results are typically a red flag" (Wilkes & Karimi, 2024, p. 9). We have already discussed how level-1 variance is not expected to change upon addition of level-2 variables. The "dramatic" change in variance at level-2 is unsurprising, and simply reflecting the fact that, for many outcomes, inequality patterns are primarily additive. For instance, if being racialized as Black entails a range of experiences of discrimination, this is often reflected in a *consistent* ("universal") pattern of self-identified Black respondents having worse health outcomes than white respondents, across a range of other social identities. To the extent that there are deviations from that typical story for a given health outcome, we will see that reflected either in other additive contributions (e.g., gender, income) or, sometimes, from an interaction effect (where the outcome is particularly high or low in a stratum, contrary to what we might expect based purely on the additive components). As seen in Table 1, the PCV—which captures the extent to which level-2 additive ME variables account for inequality patterns across strata—is typically large, usually >80% and frequently >90%.

Far from being a "red flag," all of this is entirely expected, and none of it disqualifies MAIHDA as a valid approach. Indeed, the ability to estimate these values is a key advantage of the MAIHDA approach over the single-level approach advocated by WK.

3.4. Clarification #4 – What are the goals of MAIHDA analysis?

Moving beyond misunderstandings of the model specification and interpretation, WK ask: "What does the MAIHDA method explain?" This is a reasonable question, though we would rephrase it as "What are the *goals* of MAIHDA?" Importantly, MAIHDA itself is not intended to *explain* anything, in that it is not an approach for estimating causal effects. Intersectional MAIHDA has been explicitly presented as a theoretically-driven *descriptive* approach (Evans, 2019b; Evans et al., 2023, 2024), which we hope will be applied to advance, among other things, the critical and transformative aims of Black feminist scholars who first developed intersectionality (Cho et al., 2013; Collins, 1990, 2009; Crenshaw, 1989). Intersectional MAIHDA helps to quantify, descriptively, patterns of inequality across populations defined by the researcher. The approach remains, so far, agnostic about any causal explanations that may account for those patterns.

The “theoretically-driven” part of the approach comes into play when we encourage scholars to explicitly engage with critical theories (e.g., intersectionality, critical race theory, queer theory, and other anti-oppressive frameworks) and theories of the social determinants of health (or comparable theories in adjacent fields). These theories should guide both the selection of research questions (and thus, definitions of strata) and interpretations of observed differences. Thus, what are the goals of the MAIHDA approach? There are many, including: to quantify inequalities using an intersectional framework, to facilitate inclusion of more strata—and thus more diversity—in the analysis, to better leverage available data (even when sample sizes are small), and to obtain more robust, reliable estimates. More generally, the use of MAIHDA supports critical analysis and transformative aims.

MAIHDA enables us to conceptualize and quantify inequalities in multiple ways (Evans et al., 2020)—to predict means and mean differences, and to estimate variation within and between strata (thus avoiding the “tyranny of the averages problem,” whereby decision making is based solely on group averages with no regard to the often substantial individual heterogeneity around those averages (Merlo & Wagner, 2013)). The approach enables us to consider inequality in both a broad sense—describing patterns across all strata—and to focus on specific strata or stratum comparisons. It accomplishes the former by providing statistics such as the VPC and the PCV, which describe general patterns rather than reference specific strata. The latter is addressed by enabling the prediction of stratum-specific means.

WK critique the separation of additive and interaction effects in MAIHDA (“splitting the multiplicative into additive and interactive components” (Wilkes & Karimi, 2024, p. 5)), quoting Block et al. (2023) who say “it does not make conceptual or theoretical sense to claim that the effects of things like gender and race can be broken up into separate “additive” and “interactive” or “intersectional” components.” This is an odd argument, since WK are in favor of using a single-level model with interactions, which also separates additive and interaction effects, albeit in a different way than MAIHDA. In any case, we have two responses to this critique.

First, to the argument that it does not make theoretical or conceptual sense to try to partition the *effects* of gender and race (in a causal sense) from each other and across additive and interaction effects—we actually agree. However, it is a misunderstanding of MAIHDA to assume that this is the purpose of the additive ME model, and suggesting this amounts to a strawman argument. The confusion here may reside with the use of words like “explain” or “effects”—both of which can evoke causal imaginings. For instance, WK describe how MAIHDA supposedly engages in “a tautological logic in which the individual-level demographic variables are used to create the strata and then are again used to explain why the strata are compositionally different” (Wilkes and Karimi 2024:3, underline added for emphasis). However, when discussing components of regression models, the same words (“effect” and “explain”) are used with different meanings.

For example, when we consider the extent to which additive versus interaction parameters “explain” inequalities, or whether something is an additive versus interaction “effect,” we do not actually mean to partition attribution of *causation* by gender or race/ethnicity across additive or interaction parameters. Such an understanding is antithetical to the entire intersectional approach we are using, which posits that we cannot understand these axes separately. When discussing models, by “explain” we simply mean that the variance is accounted for by, or loads onto, a parameter included in the model. A fixed (additive) or random (interactive) “effect” is distinguishing the part of the model it belongs to. This is not necessarily true in a causal sense. The uses of the additive ME model, as we have discussed, includes a descriptive evaluation of the universality-versus-specificity of inequality patterns, and improved estimation of outcomes in strata with small sample sizes. Similarly, Block et al. (2023) advocate the use of interaction models for similar purposes (although they only discuss single-level models, not MAIHDA, and therefore do not compare the approaches).

Where we differ from Block et al.'s assessment is when they conclude that in situations where interaction effects are zero, then "we would have to conclude that there is no empirical support for intersectionality" (Block et al., 2023, p. 801). Such an interpretation is common enough in the literature, and some take this further to argue that it is not worth continuing to do an intersectional analysis when the observed inequality patterns are purely (or mostly) additive. We disagree. We feel this falls into the problematic "intersectionality as testable hypothesis" treatment of the approach that Hancock (2013) has warned of. Intersectionality is best thought of as a framework for analysis and critique, rather than a testable (falsifiable) theory. It guides attention to systems of marginalization and oppression, and how they are interlocking and inseparable. "Failure" to find statistically significant interaction effects for a given outcome in a given population at a particular, observed time does not disprove intersectionality. This simply means that the inequalities produced by the social determinants associated with those systems of oppression may follow more predictable, or universal patterns, and that interaction parameters are not needed to describe the extent to which particular strata deviate from those patterns in specific or unique ways.

Second, on the issue of separating additive and interaction effects, we agree with those who argue that, on their own, interaction effects (u_j in the ME model) are difficult to interpret without also considering the magnitude of associated additive effects. Indeed, we encourage researchers to interpret specific values of u_j alongside the model's additive main effects for a fuller understanding of their meaning. Furthermore, prediction of stratum means should always be based on a (re)combination of additive main effects and stratum-level random effects.

Part of the confusion about the additive main effects of the MAIHDA model may stem from uncertainty about the uses of this model and the interpretation of the fixed effects portion on its own (separate from the stratum random effect). While this model partitions inequality (variance) across fixed additive parameters and random interaction terms, we would not generate a final predicted value for strata based only on the additive ME. As described above, the additive components of the model capture the degree to which the between-stratum inequalities are *universal*. For instance, comparing any two stratum identities (e.g., women versus men), we might ask if the inequality ‘gap’ in predicted values between them is *universal* across other axes of comparison (e.g., whether the straight women versus straight men inequality == the bisexual women versus bisexual men inequality). The additive fixed effects collectively describe the best estimate of the inequality patterns in a world where these inequalities are assumed to be (or constrained to be) the same. Therefore, the strata residuals represent deviations from those universal patterns—or the extent to which one must be *specific* about which strata (or comparisons) we are focusing on. Moreover, individual strata are allowed to distinguish themselves from the general patterns independent of other strata.

Predictions for strata based on additive main effects alone are therefore an entirely hypothetical construct, where $u_j = 0$ in each stratum. Under this hypothetical null hypothesis, where there is no specificity required to account for inequality patterns between strata, the additive main effects provide estimates of those inequalities. If the PCV were to equal 100% (and therefore all strata had residuals where $u_j = 0$), then we would be unable to reject this null hypothesis, and we would conclude that the observed inequalities are describable with consistent, universal patterns. However, if some or all strata have residuals where $u_j \neq 0$, then we should approach descriptions of inequalities with more specificity about which strata we are considering, rather than relying on universal statements.

The partitioning of variance across additive and interaction parameters, and calculation of the PCV, thus enables an exploration of the universality-versus-specificity of observed inequality patterns, bridging the consideration of the broad/general with the specific/unique. However, we reiterate that this remains only one of MAIHDA’s potential goals.

4. Conclusion

Far from being an example of a method that "prioritize[s] technical sophistication over theoretical-methodological rigor" (Wilkes & Karimi, 2024, p. 3), MAIHDA is, at its heart, a fairly straightforward approach for accomplishing what scholars in the health and social sciences have demanded—the ability to quantify inequalities across populations, to conceptualize inequalities both in terms of averages and variances, and to include more axes of marginalization, with more points of intersection made visible, despite the common occurrence of small sample size issues. By leveraging the simplicity and elegance of the well-established multilevel modeling framework, with its practical methodological advantages, and avoiding unnecessary addition of numerous fixed interaction parameters for high-dimensional analysis, MAIHDA accomplishes the opposite of what its critics claim—used correctly, it *enables* theoretical and methodological rigor.

MAIHDA has been validated through simulation studies and in empirical applications by multiple, independent research teams, and consistently compares favorably with alternative approaches, including the single-level approach advocated by WK (Bell et al., 2019; Evans et al., 2018, 2020; Mahendran et al., 2022b, 2022a). We have shown that the concerns raised in the recent critique are unfounded and are based on misreadings of the literature and misunderstandings of how MAIHDA models are specified and interpreted.

MAIHDA remains "the new gold standard" (Merlo, 2018) for estimating inequalities in an intersectional framework. Recent extensions of MAIHDA for use with random slopes/coefficients (Evans et al., 2023) have unlocked a variety of study designs and research questions, such as longitudinal analysis with MAIHDA (Bell et al., 2023). Recent publications hint at MAIHDA's as as-yet untapped potential in social epidemiology and related fields, such as clinical and medical research (Evans, 2024), education (Keller et al., 2023; Prior et al., 2022; Prior & Leckie, 2023), and environmental justice (Alvarez et al., 2022; Alvarez & Evans, 2021). We encourage scholars to familiarize themselves with this approach and consider its use in their research. Ultimately, it is our hope that MAIHDA will inspire widespread, self-reflective, and inclusive scholarship based on critical theory, and that these applications will motivate transformational work in pursuit of equity and justice.

References

Alvarez, C. H., Calasanti, A., Evans, C. R., & Ard, K. (2022). Intersectional inequalities in industrial air toxics exposure in the United States. *Health and Place*, *77*. https://doi.org/10.1016/j.healthplace.2022.102886

Alvarez, C. H., & Evans, C. R. (2021). Intersectional environmental justice and population health inequalities: A novel approach. *Social Science and Medicine*, *269*. https://doi.org/10.1016/j.socscimed.2020.113559

Axelsson Fisk, S., Mulinari, S., Wemrell, M., Leckie, G., Perez Vicente, R., & Merlo, J. (2018). Chronic Obstructive Pulmonary Disease in Sweden: An intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy. *SSM - Population Health*, *4*, 334–346. https://doi.org/10.1016/j.ssmph.2018.03.005

Barrie Wetherill, G., Duncombe, P., Kenward, M., Köllerström, J., Paul, S., & Vowden, B. (1986). Multicollinearity. In *Regression Analysis with Applications. Monographs on Statistics and Applied Probability*. Springer. https://doi.org/https://doi.org/10.1007/978-94-009-4105-2_4

Bauer, G. R., Churchill, S. M., Mahendran, M., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM - Population Health*, *14*. https://doi.org/10.1016/j.ssmph.2021.100798

Bell, A., Evans, C., Holman, D., & Leckie, G. (2023). Extending intersectional Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) for longitudinal data, with application to mental health trajectories in the UK. *Center for Open Science*, *No. jq57s*.

Bell, A., Holman, D., & Jones, K. (2019). Using shrinkage in multilevel models to understand intersectionality: a simulation study and a guide for best practice. *Methodology*, *15*(2), 88–96. https://doi.org/https://doi.org/10.1027/1614-2241/a000167

Block, R., Golder, M., & Golder, S. N. (2023). Evaluating Claims of Intersectionality. *Journal of Politics*, *85*(3), 795–811. https://doi.org/10.1086/723813

Brubaker, R. (2002). Ethnicity without groups. *European Journal of Sociology*, *43*(2), 163–189.

Cho, S., Crenshaw, K., & McCall, L. (2013). Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis. *Signs: Journal of Women in Culture and Society785*, *38*(4), 785–810.

Collins, P. H. (1990). Black Feminist Thought in the Matrix of Domination. *Collins, Patricia Hill. "Black Feminist Thought in the Matrix of Domination." Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, *138*, 221–238.

Collins, P. H. (2009). *Black Feminist Thought*. Routledge.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *U. Chi Legal F*, 139–167.

Cubells, J., Miralles-Guasch, C., & Marquet, O. (2024). Traffic pollution as a privilege: An intersectional approach to environmental justice and transport emissions. *Transportation*

Research Part D: Transport and Environment, *126*, 104032. https://doi.org/10.1016/j.trd.2023.104032

Evans, C. R. (2015). *Innovative Approaches to Investigating Social Determinants of Health-Social Networks, Environmental Effects and Intersectionality* [ScD Dissertation, Harvard T.H. Chan School of Public Health]. http://nrs.harvard.edu/urn-3:HUL.InstRepos:23205168

Evans, C. R. (2019a). Adding interactions to models of intersectional health inequalities: Comparing multilevel and conventional methods. *Social Science & Medicine*, *221*, 95–105. https://doi.org/10.1016/j.socscimed.2018.11.036

Evans, C. R. (2019b). Modeling the intersectionality of processes in the social production of health inequalities. *Social Science & Medicine*, *226*, 249–253. https://doi.org/10.1016/j.socscimed.2019.01.017

Evans, C. R. (2019c). Reintegrating contexts into quantitative intersectional analyses of health inequalities. *Health and Place*, *60*. https://doi.org/10.1016/j.healthplace.2019.102214

Evans, C. R. (2024). Overcoming combination fatigue: Addressing high-dimensional effect measure modification and interaction in clinical, biomedical, and epidemiologic research using multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). *Social Science & Medicine*, *340*, 116493. https://doi.org/10.1016/j.socscimed.2023.116493

Evans, C. R., Leckie, G., & Merlo, J. (2020). Multilevel versus single-level regression for the analysis of multilevel information: The case of quantitative intersectional analysis. *Social Science and Medicine*, *245*. https://doi.org/10.1016/j.socscimed.2019.112499

Evans, C. R., Leckie, G., Subramanian, S. V., Bell, A., & Merlo, J. (2024). A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). *SSM - Population Health*, 101664. https://doi.org/10.1016/j.ssmph.2024.101664

Evans, C. R., Nieves, C. I., Erickson, N., & Borrell, L. N. (2023). Intersectional inequities in the birthweight gap between twin and singleton births: A random effects MAIHDA analysis of 2012–2018 New York City Birth data. *Social Science & Medicine*, *331*, 116063. https://doi.org/10.1016/j.socscimed.2023.116063

Evans, C. R., Williams, D. R., Onnela, J. P., & Subramanian, S. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science and Medicine*, *203*, 64–73. https://doi.org/10.1016/j.socscimed.2017.11.011

Hancock, A.-M. (2013). Empirical Intersectionality: A Tale of Two Approaches. *3 UC Irvine L. Rev*, *259*. http://muse.jhu.edu/journals/theory_and_event/v015/15.3.hancock.html.

Holman, D., Salway, S., & Bell, A. (2020). Mapping intersectional inequalities in biomarkers of healthy ageing and chronic disease in older English adults. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-69934-8

Jones, K., Johnston, R., & Manley, D. (2016). Uncovering interactions in multivariate contingency tables: a multi-level modelling exploratory approach. *Methodological Innovations*, *9*. https://doi.org/10.1177/2059799116672874

Keller, L., Lüdtke, O., Preckel, F., & Brunner, M. (2023). Educational Inequalities at the Intersection of Multiple Social Categories: An Introduction and Systematic Review of the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA)

Approach. *Educational Psychology Review*, *35*(1), 31. https://doi.org/10.1007/s10648-023-09733-5
Krieger, N. (2011). *Epidemiology and the People's Health: Theory and Context*. Oxford University Press.
Lizotte, D. J., Mahendran, M., Churchill, S. M., & Bauer, G. R. (2020). Math versus meaning in MAIHDA: A commentary on multilevel statistical models for quantitative intersectionality. In *Social Science and Medicine* (Vol. 245). Elsevier Ltd. https://doi.org/10.1016/j.socscimed.2019.112500
Mahendran, M., Lizotte, D., & Bauer, G. R. (2022a). Describing Intersectional Health Outcomes: An Evaluation of Data Analysis Methods. *Epidemiology*, *33*(3), 395–405. https://doi.org/10.1097/EDE.0000000000001466
Mahendran, M., Lizotte, D., & Bauer, G. R. (2022b). Quantitative methods for descriptive intersectional analysis with binary health outcomes. *SSM - Population Health*, *17*. https://doi.org/10.1016/j.ssmph.2022.101032
Merlo, J. (2018). Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. *Social Science and Medicine*, *203*, 74–80. https://doi.org/10.1016/j.socscimed.2017.12.026
Merlo, J., & Wagner, P. (2013). The tyranny of the averages and the indiscriminate use of risk factors in public health: a call for revolution. *European Journal of Epidemiology*, *28*(1, Supplement), 148. https://doi.org/10.1007/s10654-013-9820-0
Moreno-Agostino, D., Woodhead, C., Ploubidis, G. B., & Das-Munshi, J. (2023). A quantitative approach to the intersectional study of mental health inequalities during the COVID-19 pandemic in UK young adults. *Social Psychiatry and Psychiatric Epidemiology*. https://doi.org/10.1007/s00127-023-02424-0
Phelan, J. C., & Link, B. G. (2015). Is Racism a Fundamental Cause of Inequalities in Health? *Annual Review of Sociology*, *41*, 311–330. https://doi.org/10.1146/annurev-soc-073014-112305
Prior, L., Evans, C., Merlo, J., & Leckie, G. (2022). *Sociodemographic inequalities in student achievement: an intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) with application to students in London, England*.
Prior, L., & Leckie, G. (2023). *An investigation of student intersectional sociodemographic and school variation in GCSE final grades in England in 2020* (Bristol Working Papers in Education Series. Working Paper: 03/2023).
Subramanian, S. V., & O'Malley, A. J. (2010). Modeling neighborhood effects: The futility of comparing mixed and marginal approaches. In *Epidemiology* (Vol. 21, Issue 4, pp. 475–478). Lippincott Williams and Wilkins. https://doi.org/10.1097/EDE.0b013e3181d74a71
Wilkes, R., & Karimi, A. (2024). What does the MAIHDA method explain? *Social Science & Medicine*, *345*, 116495. https://doi.org/10.1016/j.socscimed.2023.116495
Yang, K. (2023). Loneliness and intersectionality: A progressive conditional approach. *Current Research in Behavioral Sciences*, *5*. https://doi.org/10.1016/j.crbeha.2023.100122

Table 1. Level-1 and Level-2 Variances, Variance Partition Coefficients (VPCs) in Null Models, and Proportion change in Variance (PCV) statistics reported in recent studies using Intersectional MAIHDA

			Null Model		Main Effects Model			
Publication	Outcome *	Linear or Logistic	Level-1 Variance	Level-2 Variance	Level-1 Variance	Level-2 Variance	VPC (Null Model)	PCV
(Zettermark et al., 2021)	Association between Hormonal Contraception and Antidepressant use	Logistic	—	0.29	—	0.02	0.5%	94.5%
(Nieves et al., 2023)	Birthweight	Linear	279638	8747.2	279640	1573.1	3.0%	82.0%
(Evans, 2019)	Marijuana use**	Logistic	—	NR	—	NR	4.0%	89.5%
(O'Sullivan et al., 2023)	Type 2 Diabetes (high blood sugar)	Logistic	—	NR	—	NR	4.3%	92.3%
(Evans et al., 2018)	Body Mass Index (BMI)	Linear	34.51	1.82	34.51	0.64	5.0%	64.7%
(Evans & Erickson, 2019)	Depression (CES-D)	Linear	60.68	3.44	60.03	0.17	5.3%	94.7%
(Zubizarreta et al., 2022)	Human papillomavirus Vaccination	Logistic	—	0.2	—	0.03	5.5%	84.3%
(Persmark et al., 2019)	Prescription Opioid Misuse	Logistic	—	0.268	—	0.016	7.5%	94.1%
(Balloo et al., 2022)	Mental health (mental distress)***	Logistic	—	0.31	—	0.01	8.6%	97.8%
(Moreno-Agostino et al., 2023)	Mental health (depressive symptomatology)	Linear	NR	NR	NR	NR	9.2%	95.5%
(Evans, 2019)	Cigarette use**	Logistic	—	NR	—	NR	11.5%	88.8%
(Axelsson Fisk et al., 2018)	Chronic Obstructive Pulmonary Disease	Logistic	—	0.83	—	0.04	20.0%	95.2%
(Beccia et al., 2021)	Eating-related pathology (lifetime eating disorder diagnosis)	Logistic	—	1.3	—	0.1	28.5%	93.1%

Notes: Table rows are sorted from low to high VPC reported for null models. Papers were selected in a non-systematic way to provide a sense of the different ways the MAIHDA approach has been used in publications, and to demonstrate a range of VPC and PCV values observed. VPC and PCV proportions are converted to percentages, as is typical in the literature. Logistic models do not estimate Level-1 variances, and so will not report them. The VPC in such cases is calculated using a latent response approach, where the level-1 variance is set equal to $\pi^2/3 \approx 3.29$ where π is the mathematical constant 3.142.
NR = not reported in article tables.

* Some studies examined multiple outcomes. Unless otherwise noted, results are presented only for the outcome specified in parentheses. For Zettermark et al. (2021) and Beccia et al. (2021), we present the outcomes discussed by WK.
** Study examined many outcomes with various definitions of strata. Previous 30-day cigarette/marijuana use (yes/no) with 91 strata is shown.
*** Results presented for respondents who attended university.