SUPPLEMENTARY MATERIAL

This document is divided into several sections. The first section describes the methodology used to identify, gather, verify, organize, and evaluate the studies that form part of the samples used in our meta-analysis. The second section is about sensitivity analysis and includes multiple tests implemented to ensure our results are robust to known validity threats. The third section contains tables and plots that we could not include in the main body of the paper.

**I. Data collection**

1. PRISMA

The last column of Table S1 identifies descriptions of PRISMA items and the location in the paper. The review protocol was not registered in PROSPERO, as the authors did not know the database before the data extraction was initiated. However, a search in Prospero database was carried out to investigate whether similar reviews had already been registered. By October 5, 2023, a search was made searching for terms “migrant health” and found 51 results. A thorough review of these protocols showed that with one exception all reviews focused on migrant health without contrasts with host populations/peers-in-origin and just one review protocol partially overlapped with the meta-analysis targeted populations. The exception is a review

(<http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42023445035>)

that was limited to Middle East and North African migrants and, furthermore, was registered as a review protocol in July, 2023.

2. Database for analysis

2.1. Identification of studies, inclusion and exclusion rules

We searched PUBMED and SCOPUS databases for studies published prior to May 2021, reporting health outcomes associated (and some not) with DOHaD in migrants, without language restrictions. Full-search terms and complete inclusion and exclusion criteria are described in the Tables S2. We further searched reference lists and relevant reviews to identify additional studies of interest. We screened titles and abstracts identified by the original database search and independently examined citations and reference lists to generate a comprehensive bibliography relevant to our review. We used an agreed upon coding form and note extraction program for Zotero[[1]](#footnote-1) to extract information for each outcome and migrant group reported in studies that met the inclusion criteria. (see below 2.2). As an additional step, excluded studies were reviewed by a second researcher to ensure all of them fully met one or more of the exclusion criteria. Table S4 includes a list of the studies that were finally excluded due to the age range (adolescents) and thus could not address our hypotheses even though their design fitted our inclusion criteria. We removed duplicate data for studies reporting information from the same migrant group sample, for the same outcome, and the same period, keeping the most complete report.

2.2. Data evaluation and organization

The data thus extracted was cleaned and organized in spread sheets. We evaluated the quality of each study included in the database using a variant of the Newcastle-Ottawa scale. Description of the scale is in Supplementary Text S1 and the final result of the quality assessment evaluation is presented in Table S5.

2.3. Information retrieved from each study

We included the following items:

i. Publication details (author, year, outlet);

ii. Study design and study analytics including: country/region of native and migrant populations, sampling information (type of sample, sample sizes), nature of method(s) for the assessment of outcomes’ prevalence, covariates employed in model estimation, class of model used to estimate effects)

iii. Key results in the form of estimated effects on prevalence ratio or log-odds, standard errors and confidence intervals)

2.4. Classification of studies

Table S3 contains a narrative synthesis of the included studies and a classification of these studies according to (i) target health outcome, (ii) migrant region of origin and (iii) key results.

2.5. Outcome standardization

To calculate a single pooled effect per outcome, we transformed some of the extracted original effects to ensure that the same reference group was used in all results reported in the study. In studies reporting estimates of prevalence in the reference group but only estimated effects on odds ratios, we employed Zhang and Yu’s (1998) approximation to compute effects on prevalence. While this transformation is not without its limitations, it may be better than interpreting ORs as a measure of relative risk in cases in which the outcome of interest is not rare in the population. This enabled us to pool studies that only reported odds ratios with those that focused on prevalence. Six studies did not include enough information to transform the effects into prevalence ratios, so we pooled their estimated effects as odds ratio. These studies are identified in Table 1 (see also Table S3).

We also tested if the inclusion/exclusion of these studies along with those that reduced the host population to native ethnic counterparts or minority populations affected the results of our analysis. Table S9 is a replica of Table 2 reporting the main results of our study, but excluding those articles with untransformed odds ratios and including those that restrict the host population to a native minority.

**II. Sensitivity analysis.**

There are a several threats to validity of the foregoing findings and potential interpretational difficulties. These include (in no special order), publication bias, clustering effects, unbalanced study weights, and between study heterogeneity. To minimize these threats, we performed multiple tests to detect potential flaws and assess the magnitude of their effects on inferences. Because our strongest findings are those associated with T2D, we show mostly results for that outcome although, in some cases and for illustration purposes only, we also display results for the other outcomes.

A. Adjustments for multiple observations for one outcome retrieved from a single study.

For all outcomes we study here, a single study may contribute multiple observations. For example, study Y provides estimates of effect sizes for, say T2D, and it does so using different contrast groups, different age groups, different survey waves etc. Some of these traits we can control for either via subgroup analysis or using moderators. Since to control for all those that appear relevant is costly in terms of degrees of freedom, most of our analyses are based on estimated effects retrieved from the same study without controlling all relevant characteristics. This leads to clustering of observations derived from the same study that produces two problems. First, repeated estimates for the same outcome from a single study potentially creates “clustering effects” of studies analogous to clustering of correlated observations. As a consequence, standard errors of effects will be underestimated. Second, if a single study contributes with many more observations to a single outcome than the others, estimates could be excessively influenced by results from that single study. To reduce both problems we implement a bootstrap-like procedure as follows.

i. For each outcome, we organize observations using the study contributing to it. We create groups containing one or more observations contributed by the same study. If the number of observations in a group exceeds 5, we create a cluster, say, C. Let NC be the number of observations in cluster C. All observations retrieved from a study that contribute with less than 5 observations were included in a residual cluster, RC with effective size NRC

ii. We randomly draw one observation per cluster and proceed to estimate models in this much reduced sample of studies representing only one observation of the outcome per study. We store the estimates and its standard error.

iii.. The previous step is repeated 100 times and we collected 100 replicas of estimates of effect sizes and corresponding standard errors that are unaffected by clustering

iv. We compute the mean of the replicas and the estimated standard of the mean using Rubin’s rule. We then use these statistics to test the null hypotheses that the mean effect is 0. Results of this exercise for T2D are in Figure S1. The key inference is that even though the standard error from the RE is indeed underestimated by a factor of 2-3, a new statistical test that uses the mean and adjusted standard error from 100 simulations, will continue to reject the null hypotheses that the effect size of migrant status on T2D is 0., albeit with a slightly wider 95% CI. This inference is unaffected by clustering effects or disparities in the weight that each study exerts on the analysis.

B. Meta-analytic robustness tests

The following tests, some native to meta-analysis and some specially designed by us for this study, focus on three issues, unexplained study heterogeneity, small-study bias (or study outliers), and publication bias.

*B.1. Heterogeneity*

Between-study variability of estimates of effect sizes is reflected in non-overlapping CI’s associated with the studies that entered in the estimation. This can be graphically seen in forests plots in the last section of this supplement. It is a result of differences in study design and data collection strategies (including differences in survey questionnaire) and/or of differences in target populations, definitions of outcomes and of covariates.

There are several statistics that either detect the presence of lack of homogeneity (Q statistic) or assess its magnitude ( and ). The Q and (and associated significance tests) are displayed in the tables with results in the main body of the paper. However, it is also useful to explore other strategies to gain more insight about the potential nature and effects between study variability. As mentioned in the body of the paper, we use I2 as just one indicator of heterogeneity but complement this via inspection of the range of effect size estimates. We combine these two ways of assessing heterogeneity as a defense against relying too heavily on potential flaws inherent in I2. Most of our findings are based on estimates with heterogeneity reinforcing the conjectures as they all have the same sign and center around values quite removed from the no-effect value (1).

i. Multilevel model:

One way to identify the source of between-study heterogeneity is to estimate a multilevel (three-level) meta-analysis model in which the top level are studies that contribute with more than a handful of observation to a given outcome and the lower level is the observation itself. The multi-level approach addresses the potential confounding of the estimated pooled effects that could be generated by the inclusion of multiple observed effects from a single study. The standard random-effect model computes between-study heterogeneity as a source of variability (other than sampling error) caused by between-study variation associated with study designs and target populations. The multi-level model adds a third source of variation to account for study specific differences that could influence multiple observations of the same outcome retrieved from any one of these studies70. The multilevel model introduces a second term to account for heterogeneity effects that are within-study specific as different from heterogeneity effects that are between-studies specific. Introducing this third level enables us to gauge if and to what an extent repeated observations from a single study are exerting undue influence on the estimates of the overall effect size. The multilevel approach is only relevant for *global estimates* of effect sizes associated with each outcome, not for estimates that apply to analysis by subgroups (of origins and destination).

Results from this procedure applied to T2D are in Table S6 and the corresponding plot is in Figure S2. The effect size continues to be quite large and statistically significant (PR 1·61; 95% CI 1·27-2·03), although the confidence interval has widened due to the inclusion of a third source of variation. Most interestingly, multilevel model shows that 53% of the calculated variation from the true effect size may be attributed to differences between clusters (between studies) while 44% is associated with effect heterogeneity within (cluster of) studies. This suggests that even when the inclusion of repeated observations from same studies has modest influence on results, estimates of effects are robust to this source of confounding.

ii. Graphical representation

One of the graphical tools (there are others) is the Galbraith plot and associated statistics. The central idea is to assess the relation between the values of the standardized (z-score) estimated effect and its precision (the inverse of the standard error). Plots for all five outcomes are in Figure S3. The key feature of these plots is that in each of them the bulk of observations hug the regression line (red line) whose slope is equal to the average effect and fall within a 95 percent CI. This is an indication of moderate between-study variability. Note also that in the case of T2D and child obesity, though less so for hypertension and adult obesity, most of the departures from the 0-effect horizontal line have a positive sign, i.e. the studies point to a positive effect of migrant status on the prevalence of the condition. From these figures we conclude that between-study heterogeneity should be of some concern only in the case of adult obesity and hypertension.

iii. Moderators

To reduce between-study heterogeneity, we employed estimation of models including moderators (results not included in Supplemental materials, available on request) as well as estimation of models by subgroups, which is our preferred strategy and used throughout the paper. The key variable to define subgroups in the analysis is region of origin. We also employed more fined-tuned subgroups that were combined with region of destination. Thus, for example, we estimated models according to the combination of region of origin and destination and using different contrast groups, e.g. native versus populations of origin (see main body of paper). The remaining heterogeneity is unexplained but its implications for our inferences are relevant only when heterogeneity is accounted for by both negative and positive deviations from the overall mean effect. In our case the bulk of heterogeneity is produced by variability in the “right” direction, namely, most is due to between-study differences of estimates of effects that have the same sign. At least in the case of adult and child obesity and T2D, the between-study heterogeneity that remains after using either moderators or subgroups consists of between-studies variability estimates that lead to the same inferences, i.e. have the same sign. It is less so in the case of hypertension.

*B.2 Study outliers*

It is possible that any of the analyses we carry out are influenced by studies with large (small) effects that are outliers. To investigate this, we first employ an overall assessment to detect the presence of influential studies and then a more nuanced evaluation of sensitivity of results to any one study included in the sample.

i. Overall assessment

We use the so-called Egger’s test which serves to identify small studies that may have unduly large influence on estimates. We conducted this test on T2D, hypertension, and adult and child obesity. This test can also be useful to account for between-study heterogeneity as those with large effects contribute disproportionately to any indicator of inter-study heterogeneity. The results displayed in Table S7 suggest that there is no strong evidence of outlier studies in any of the outcomes we study and, in addition, the little there is vanishes after controlling for subgroups of origin.

ii. Leave-one-out analysis

A second strategy to detect the presence of outliers (and of identifying potential sources of heterogeneity) is to repeat estimation of a model by selecting out a single study from the sample. That is, if there is a group of K observations, there will be K new estimates retrieved from samples of K-1 observations each. Figures S4 display results for T2D and Obesity. If the presence of outliers were influential, one would expect that some of the K estimates should be out of line either in terms of absolute values or in terms of their CI’s. Figures S4 shows that there are no indications that this is occurring in our case.

*B.3 Publication bias*

One of the most common problems in meta-analysis is the so-called publication bias. This occurs when the observed published studies do not represent well the universe of studies that could have been found had all of them been published. What usually happens is that low-powered, small studies, that report no statistically significant estimates, are left out because they are less likely to be published in peer-reviewed journals. There are tools to detect the bias and to assess the potential size of these biases. Tables S8 and Figures S5 display results of a test identified as trim-and-fill. This test consists of first detecting publication bias via a funnel plot and then imputing possible non-published studies that would eliminate indications of publication bias detected in the funnel plot. A cursory examination of Tables S6 and Figures S5 show that there is no evidence of publication biases for T2D and hypertension, very little for child obesity, and a modest amount for adult obesity. The outcome hypertension included too few studies to perform a trim and fill test.

|  |  | Table S2: PRISMA checklist(1)  **Table S1: PRISMA checklist(1)** |  |
| --- | --- | --- | --- |
| **Section and Topic** | **Item #** | **Checklist item** | **Location where item is reported** |
| **TITLE** | | |  |
| Title | 1 | Identify the report as a systematic review. | 1 |
| **ABSTRACT** | | |  |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist. | 2 |
| **INTRODUCTION** | | |  |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge. | 3-7 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | 3-7 |
| **METHODS** | | |  |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Figure 1 and Suppl. Materials |
| Information sources | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | 7-9 and Suppl.Materials |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Suppl.Materials |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Suppl.Materials |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Suppl.Materials |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | 9 and Suppl.Materials |
| 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | 10-11 and Suppl.Materials |
| Study risk of bias assessment | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Suppl.Materials |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | 10 and Table 1 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)). | Figure 1, 7-9 and Suppl.Materials |
| 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Suppl.Materials |
| 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | 10-14 |
| 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | 10-14 |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Suppl.Materials |
| 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Suppl.Materials |
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Suppl.Materials |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | 10-14 and Suppl.Materials |
| **RESULTS** | | |  |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | 10-14, Figure 1 |
| 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Table S4 |
| Study characteristics | 17 | Cite each included study and present its characteristics. | Table 1 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study. | Suppl.Materials |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Tables 2,3a-3g and 4a-4g |
| Results of syntheses | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Suppl.Materials |
| 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Tables 2,3a-3g,4a-4g and Suppl.Materials |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Suppl.Materials |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Suppl.Materials |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Suppl.Materials |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Suppl.Materials |
| **DISCUSSION** | | |  |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence. | 14-17 |
| 23b | Discuss any limitations of the evidence included in the review. | 14-17 |
| 23c | Discuss any limitations of the review processes used. | 14-17 and Suppl.Materials |
| 23d | Discuss implications of the results for practice, policy, and future research. | 14-17 |
| **OTHER INFORMATION** | | |  |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | Suppl.Materials, 1 |
| 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | Suppl.Materials |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | 17 |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | 17 |
| Competing interests | 26 | Declare any competing interests of review authors. | 17 |
| Availability of data, code and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | 17 |

*(1) From:*  Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: <http://www.prisma-statement.org/>

Table S2: Full search terms

|  |
| --- |
| **Databases:** PUBMED and SCOPUS |
| **Publication dates:** prior to May 2021 |
| **Other limits:** None |
|  |
| **Search string: [All fields]** |
|  |
| (migrant OR immigrant OR immigration OR “ethnic minority” OR “migrant origin”) |
|  |
| AND |
|  |
| (obesity OR overweight OR “body mass index” OR BMI OR waist to hip ratio OR waist circumference OR subscapular adiposity OR adiposity OR “metabolic syndrome” OR diabetes OR T2D OR “diabetes mellitus” OR hyperglycemia OR non-alcoholic fatty liver disease (NAFLD) OR hypertension OR cognitive decline OR liver disease OR cardiovascular disease OR disability OR methylation OR acetylation OR RNA interference or epigenetics OR genetics OR DOHAD or Predictive Adaptive Response OR obesogenic OR fetal programming OR thrifty gene OR) |

Table S3: Narrative synthesis of the 38 studies included for review

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** | **Region** | **Decreased risk** | **Non-significant** | **Increased risk** |
| **Diabetes** | North America (N = 25) | **Africans** [Ref: Host non-hispanic blacks] Palarino 2021; **HIC** [Ref: Host non-hispanic blacks] Palarino 2021 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018;  **HIC** [Ref: Host] Shiue 2014\*; **Latin American** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host] Shiue 2014\*;  **Mexican** [Ref: Host] Shiue 2014\*; [Ref: Host non-hispanic blacks] Palarino 2021 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018;  **All** [Ref: Host] Shiue 2014\*; **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic white] Veenstra and Patterson 2016;  **Latin American** [Ref: European migrants] Commodore-Mensah et al. 2018 |
| Europe (N = 19) |  | **Asian** [Ref: Host] Bodewes et al. 2021; [Ref: Origin] Koochek et al. 2008;  **HIC** [Ref: Origin] Alves et al. 2015;  **Latin American** [Ref: Origin] Verstraeten et al. 2018\* | **African** [Ref: Rural origin] Agyemang et al. 2016;  **All** [Ref: Host] Shiue 2014\*; **Asian** [Ref: Host] Bennet et al. 2015, Bodewes et al. 2021, Raza et al. 2017;  **Latin American** [Ref: Host] Verstraeten et al. 2018\* |
| Other (N = 37) | **African** [Ref: Host] Motlhale and Ncayiyana 2019; **All** [Ref: Host] Motlhale and Ncayiyana 2019; **Asian** [Ref: Host] Piao et al. 2020\* | **African** [Ref: Host] Shamshirgaran et al. 2013; [Ref: Rural origin] Oyebode et al. 2015;  **Asian** [Ref: Host] Reuven et al. 2016, Shamshirgaran et al. 2013; [Ref: Rural origin] Oyebode et al. 2015;  **HIC** [Ref: Host] Shamshirgaran et al. 2013;  **Latin American** [Ref: Rural origin] Miranda et al. 2011\*;  **Mexican** [Ref: Rural origin] Oyebode et al. 2015;  **Other** [Ref: Host] Shamshirgaran et al. 2013 | **African** [Ref: Host] Reuven et al. 2016, Shamshirgaran et al. 2013, Simchoni et al. 2020; [Ref: Rural origin] Agyemang et al. 2016, Oyebode et al. 2015;  **Asian** [Ref: Host] Shamshirgaran et al. 2013; [Ref: Rural origin] Oyebode et al. 2015;  **HIC** [Ref: Host] Shamshirgaran et al. 2013;  **Other** [Ref: Host] Shamshirgaran et al. 2013 |
|  |  |  |  |  |
| **Hypertension** | North America (N = 27) | **African** [Ref: Host non-hispanic blacks] Brown et al. 2017, Palarino 2021;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018;  **HIC** [Ref: Host non-hispanic blacks] Palarino 2021;  **Latin American** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Mexican** [Ref: Host non-hispanic blacks] Palarino 2021 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Origin] Brown et al. 2017;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Latin American** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Mexican** [Ref: Host ethnic counterpart] Salinas et al. 2008 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic white] Veenstra and Patterson 2016;  **Latin American** [Ref: European migrants] Commodore-Mensah et al. 2018 |
| Europe (N = 16) |  | **Asian** [Ref: Host] Raza et al. 2017;  **HIC** [Ref: Origin] Alves et al. 2015;  **Latin American** [Ref: Host] Diemer et al. 2020 | **African** [Ref: Rural origin] van der Linden et al. 2019;  **Asian** [Ref: Origin] Koochek et al. 2008;  **Latin American** [Ref: Host] Verstraeten et al. 2018\*; [Ref: Origin] Verstraeten et al. 2018\* |
| Other (N = 22) | **African** [Ref: Host] Motlhale and Ncayiyana 2019;  **All** [Ref: Host] Motlhale and Ncayiyana 2019; **Asian** [Ref: Host] Guo et al. 2015; [Ref: Rural origin] Oyebode et al. 2015;  **HIC** [Ref: Host] Guo et al. 2015 | **African** [Ref: Host] Reuven et al. 2016; [Ref: Rural origin] Oyebode et al. 2015;  **Asian** [Ref: Host] Guo et al. 2015; [Ref: Rural origin] Oyebode et al. 2015;  **Latin American** [Ref: Rural origin] Miranda et al. 2011\*;  **Mexican** [Ref: Rural origin] Oyebode et al. 2015;  **Other** [Ref: Origin] Gibson et al. 2013 | **African** [Ref: Host] Reuven et al. 2016; [Ref: Rural origin] van der Linden et al. 2019;  **Asian** [Ref: Host] Reuven et al. 2016 |
|  |  |  |  |  |
| **Obesity/Oveweight Adults** | North America (N = 37) | **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic white] Singh and DiBari 2019;  **All** [Ref: Host] Oh et al. 2021, Singh and DiBari 2019; **HIC** [Ref: Host non-hispanic white] Singh and DiBari 2019;  **Latin America** [Ref: Host non-hispanic white] Singh and DiBari 2019 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic blacks] Palarino 2021;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018;  **HIC** [Ref: Host non-hispanic blacks] Palarino 2021;  **Latin America** [Ref: European migrants] Commodore-Mensah et al. 2018 | **African** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic white] Singh and DiBari 2019;  **Asian** [Ref: European migrants] Commodore-Mensah et al. 2018;  **Latin America** [Ref: European migrants] Commodore-Mensah et al. 2018; [Ref: Host non-hispanic white] Singh and DiBari 2019;  **Mexican** [Ref: Host non-hispanic blacks] Palarino 2021; [Ref: Host non-hispanic white] Singh and DiBari 2019;  **Other** [Ref: Host non-hispanic white] Singh and DiBari 2019 |
| Europe (N = 24) |  | **Asian** [Ref: Origin] Koochek et al. 2008;  **HIC** [Ref: Host] Lindström and Sundquist 2005; [Ref: Origin] Alves et al. 2015;  **Latin American** [Ref: Origin] Verstraeten et al. 2018\*;  **LMIC** [Ref: Host] Lindström and Sundquist 2005 | **African** [Ref: Host] Lindström and Sundquist 2005; [Ref: Rural origin] Agyemang et al. 2016, Cohen et al. 2017\*;  **Asian** [Ref: Host] Raza et al. 2017;  **HIC** [Ref: Host] Alkerwi et al. 2012, Lindström and Sundquist 2005;  **Latin American** [Ref: Host] Verstraeten et al. 2018\*;  **LMIC** [Ref: Host] Lindström and Sundquist 2005 |
| Other (N = 44) | **African** [Ref: Host] Menigoz et al. 2016, Reuven et al. 2016;  **Asian** [Ref: Host] Guo et al. 2015, Menigoz et al. 2016; [Ref: Rural origin] Oyebode et al. 2015;  **HIC** [Ref: Host] Guo et al. 2015, Menigoz et al. 2016 | **African** [Ref: Host] Menigoz et al. 2016; [Ref: Rural origin] Oyebode et al. 2015;  **All** [Ref: Origin] Menigoz et al. 2016; **Asian** [Ref: Host] Guo et al. 2015; [Ref: Rural origin] Oyebode et al. 2015;  **HIC** [Ref: Host] Menigoz et al. 2016;  **Mexican** [Ref: Rural origin] Oyebode et al. 2015;  **Other** [Ref: Host] Menigoz et al. 2016 | **African** [Ref: Rural origin] Agyemang et al. 2016, Cohen et al. 2017\*, Oyebode et al. 2015;  **All** [Ref: Origin] Menigoz et al. 2016; **Asian** [Ref: Host] Reuven et al. 2016; [Ref: Rural origin] Oyebode et al. 2015;  **Latin America** [Ref: Rural origin] Miranda et al. 2011\* |
|  |  |  |  |  |
| **Obesity/Overweight Children** | North America (N = 2) |  | **All** [Ref: Host] Jackson 2012; **Asian** [Ref: Host ethnic counterpart] Argueza 2020 |  |
| Europe (N = 18) |  | **African** [Ref: Host] Besharat Pour et al. 2014b, Labree et al. 2015;  **All** [Ref: Host] Jackson 2012, Will et al. 2005;  **Asian** [Ref: Host] Kirchengast and Schober 2006;  **HIC** [Ref: Host] Besharat Pour et al. 2014b,  **LMIC** [Ref: Host] Labree et al. 2015;  **Other** [Ref: Host] Kirchengast and Schober 2006 | **All** [Ref: Host] Besharat Pour et al. 2014, Besharat Pour et al. 2014b, Will et al. 2005;  **Asian** [Ref: Host] Besharat Pour et al. 2014b, Labree et al. 2015;  **Latin America** [Ref: Host] Besharat Pour et al. 2014b,  **Other** [Ref: Host] Kirchengast and Schober 2006 |
| Other (N = 4) |  | **HIC** [Ref: Host] Zulfiqar et al. 2019;  **LMIC** [Ref: Host] Zulfiqar et al. 2019 |  |
|  |  |  |  |  |
| **Other (Not associated with PAR\*\*)** | North America (N = 6) | **All** [Ref: Host] Oh et al. 2021, Shiue 2014 | **Asian** [Ref: Host non-hispanic white] Veenstra and Patterson 2016 |  |
| Europe (N = 4) | **All** [Ref: Host] Shiue 2014 | **All** [Ref: Host] Shiue 2014 |  |
| LMIC: low-middle income countries  HIC: high income countries (includes all Europe, Japan, South Korea, North America, Australia, Non-Hispanic Whites)  Asia includes Russia; African includes Middle East; Latin America includes South and Central America (except Mexico)  \*Studies that were not included in the final analyses because they reported OR without sufficient information for transforming to prevalence ratios  \*\* 'Other' includes outcomes not directly invoked in DOHaD as a result of exposures to early conditions. In the studies included in our data base, these outcomes are asthma (n=5), cancer (n=3) and mental disorders (n=2). | | | | |

**Table S4. Studies excluded from the meta-analysis because the age range does not fit hypotheses**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Reference** | **Dates** | **Origin** | | | **Arrival** | **N** | **Outcomes** | **Reference group (Migrant Generation)** | **Age (Mean)\*** |
| Esteban-Gonzalo L, Veiga ÓL, Regidor E, Martínez D, Marcos A, Calle ME. Immigrant Status, Acculturation and Risk of Overweight and Obesity in Adolescents Living in Madrid (Spain): The AFINOS Study. J Immigrant Minority Health. 2015 Apr;17(2):367–74. | 2007-2008 | South America and Europe | | | Europe (Spain) | 2081 | OW | Native at destination (1) | 13-17 (15) |
| Johansson-Kark M, Rasmussen F, Hjern A. Overweight among international adoptees in Sweden: a population-based study. Acta Paediatrica. 2007 Jan 2;91(7):827–32. | 1991-1995 | Asia (Thailand, South Korea, Indonesia, Sri Lanka, India), South America (Peru, Ecuador, Colombia, Chile) | | | Europe (Sweden) | 265573 | OW | Native at destination (1) | 18 |
| Maldonado L, Albrecht S. Does the Immigrant Advantage in Overweight/Obesity Persist over Time in Mexican American Youth? NHANES 1988-1994 to 2005-2014. Obesity. 2018;26(6):1057–62. | 1988-1994, 2005-2014 | Mexico | | | North America (USA) | 4720 | OW | Ethnic native counterpart at destination (1) | 4-17 (10-12) |
| Peled A, Gordon B, Twig G, Mendlovic J, Derazne E, Lisnyansky M, Raz I, Afek A. Immigration to Israel during childhood is associated with diabetes at adolescence: A study of 2.7 million adolescents. Diabetologia, 2017;60(11), 2226-2230. | 1967-2016 | Africa (North, Ethiopia), Middle East, Former USSR, HIC | | | Asia (Israel) | 2721767 | Diabetes | Ethnic native counterpart at destination (1) | 16-19 (17-18) |
| Peled A, Gordon B, Twig G, Grossman E, Matani D, Derazne E, Afek A. Hypertension and childhood migration: A nationwide study of 2.7 million adolescents. Journal of Hypertension, 2019;37(4), 702-709. | 1967-2016 | North Africa, Middle East, Former USSR, HIC | | | Asia (Israel) | 2681294 | HT | Ethnic native counterpart at destination (1) | 16-19 (17-18) |
| Singh G, Kogan M, Yu S. Disparities in Obesity and Overweight Prevalence Among US Immigrant Children and Adolescents by Generational Status. Journal of Community Health. 2009;34:271–81. | 2003 | Asia, South America, Africa, Non-Hispanic White | | | North America (USA) | 46707 | OW | Native non-hispanic white at destination (1, 2) | 10-17 (nr) |
|  | | |  |  |  |  |  |  |  |
| \* Rounded to one full year or whole number. Range of mean age when results are presented separately for males/females or different age groups.  OW: overweight (includes obesity). HT: hypertension. HIC: High income countries. Nr: not reported. | | | | | | | | | |

**Supplementary Text S1: Quality Assessment Coding Criteria**

To evaluate each study for inclusion in the database the Newcastle-Ottawa scale was used with additional criteria included in Li and Lumey (2022) and West et al. (2023), when applicable. The evaluation’s dimensions and criteria for rankings are defined below. Two different scoring scales were used. Both scales were used by two of the authors independently and the corresponding scores were cross checked. In the first, the maximum for each dimension score is 12 and the minimum 0. In the second scale, the criteria defined below were used. We assigned a maximum of 2 points per domain if quality was *good*, 1 point if quality was *fair*, and 0 points if the quality was considered *poor*. We then added the scores assigned to each dimension to produce the total score for a single study. Because there was complete agreement between the two scales, Table S5 only shows results obtained from the simplified scale. Most of the studies in the sample scored in the ‘moderate’ range. By and large, all the studies included apply suitable statistical analysis, use large samples, include proper controls and use clear and unambiguous definitions of contrast groups (migrants and non-migrants).

Definition of dimensions utilized to evaluate studies

1. Sampling sources

*1.1. Good*: the sample groups are well-defined and represent a national population in both the country of origin and destination and the subpopulation of migrants residing in a place of destination. This criteria highlights the optimal design for our ‘quasi-experimental natural conditions’

*1.2. Fair*: the sample groups are well-defined and represent a national population in the country of destination and in the subpopulation of migrants but not the national or subnational population of origin

*1.3. Poor*: Sample of migrants residing in country of destination but sample not representative of national population at origin/destination

2. Sample size

*2.1. Good*: samples sizes of comparison groups are larger or equal to 200 (corresponding to {alpha=.05, beta0=.80, population prevalence in native group of .10 and a difference of .10}

*2.2. Fair*: sample sizes of comparison groups are larger or equal to 62 (corresponding to {alpha=.05, beta0=.80, population prevalence in native group of .10 and a difference of .20}

*2.3. Poor*: sample sizes lower than the minimum those specified above.

The above criteria are somewhat stacked *against* the alternative hypothesis of differences between native and migrant groups. If the ratio of prevalence rate falls below the closed interval [2,3] the minimum sample size will not be enough to reject the null hypotheses of equality between groups. In this case, we will not be able to confirm the DOHaD conjecture.

3. Identification of comparisons groups

*3.1. Good*: well-defined native population at destination and at origin. This means that country of birth of comparisons groups are known with precision.

*3.2. Fair*: migrant group is well defined. Ambiguous definition of host population (may include subpopulations not born in the country of interest but not in country of destination or similar)

*3.3. Poor*: ill-defined migrant population

4. Controls for relevant covariates

*4.1. Good*: includes indicators of education attainment (or SES), age, gender.

*4.2. Fair*: missing gender but not age or education (or SES)

*4.3. Poor*: missing education attainment (or SES) and age

5. Nature of outcome

*6.1 Good*: outcomes assessed via biomarkers (T2D, hypertension) or anthropometry (waist-to-hip ratio, clinician assessment of adiposity)

*6.2 Fair*: outcomes assessed via self-reports or anthropometry (BMI). Overweight included when defining obesity. Unspecified type of diabetes (T1D-T2D).

*6.3. Poor*: none of the above.

6. Quality of statistical analysis

*5.1. Good*: proper use of logit models for dichotomous or polytomous outcomes.

*5.2. Fair*: estimation of linear probability models.

*5.3. Poor*: none of the above.

**Table S5. Results of risk of bias analysis (scale 0-2)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **STUDY** | **Sampling sources** | **Sample size** | **Contrast groups** | **Covariates** | **Outcome** | **Statistics** | **Final score** |
| Agyemang 2016 | 2 | 2 | 2 | 2 | 2 | 2 | **12** |
| Alkerwi 2012 | 1 | 1 | 2 | 2 | 1 | 2 | **9** |
| Alves 2015 | 1 | 2 | 1 | 2 | 2 | 2 | **10** |
| Argueza 2020 | 1 | 1 | 2 | 2 | 1 | 2 | **9** |
| Bennet 2015 | 1 | 2 | 2 | 2 | 2 | 2 | **11** |
| Besharat-Pour 2014a | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Besharat Pour 2014b | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Bodewes 2021 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Brown 2017 | 1 | 1 | 1 | 2 | 2 | 2 | **9** |
| Cohen 2020 | 2 | 1 | 2 | 1 | 1 | 2 | **9** |
| Commodore-Mensah 2018 | 1 | 2 | 0 | 2 | 1 | 2 | **8** |
| Diemer 2020 | 2 | 2 | 1 | 2 | 2 | 2 | **11** |
| Gibson 2013 | 1 | 1 | 2 | 2 | 2 | 2 | **10** |
| Guo 2015 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Jackson 2012 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Kirchengast and Schober 2006 | 1 | 1 | 2 | 2 | 1 | 2 | **9** |
| Koocheck 2008 | 1 | 0 | 2 | 2 | 1 | 2 | **8** |
| Labree 2015 | 1 | 1 | 1 | 2 | 1 | 2 | **8** |
| Lindström and Sundquist 2005 | 1 | 0 | 1 | 2 | 1 | 2 | **7** |
| Menigoz 2016 | 1 | 1 | 2 | 2 | 1 | 2 | **9** |
| Miranda 2011 | 1 | 1 | 1 | 2 | 1 | 2 | **8** |
| Motlhale and Ncayiyana 2019 | 0 | 2 | 1 | 2 | 1 | 2 | **8** |
| Oh 2021 | 0 | 2 | 1 | 2 | 2 | 2 | **9** |
| Oyebode 2015 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Palarino 2021 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Piao 2020 | 1 | 2 | 2 | 2 | 1 | 2 | **10** |
| Raza 2017 | 0 | 1 | 1 | 2 | 1 | 2 | **7** |
| Reuven 2016 | 0 | 2 | 1 | 2 | 1 | 2 | **8** |
| Salinas 2008 | 1 | 2 | 1 | 2 | 2 | 2 | **10** |
| Shamshirgaran 2013 | 1 | 2 | 2 | 2 | 2 | 2 | **11** |
| Shiue 2014 | 1 | 1 | 1 | 2 | 1 | 2 | **8** |
| Simchoni 2020 | 0 | 2 | 1 | 2 | 2 | 2 | **9** |
| Singh and DiBari 2019 | 0 | 2 | 1 | 2 | 1 | 2 | **8** |
| van der Linden 2019 | 2 | 2 | 2 | 2 | 2 | 2 | **12** |
| Veenstra and Patterson 2016 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |
| Verstraeten 2018 | 2 | 2 | 2 | 2 | 1 | 2 | **11** |
| Will 2005 | 0 | 0 | 1 | 1 | 1 | 2 | **5** |
| Zulfiqar 2019 | 1 | 2 | 1 | 2 | 1 | 2 | **9** |

Figure S1: Estimates of effects and SE’s from 100 simulations





Table S6: Multi-level meta-analysis for T2D

|  |  |
| --- | --- |
|  | T2D |
| STATISTIC |  |
| Theta | .48 |
| 95% CI | [0.24;0.71] |
| Z  Prob> | 3.99  0.0001 |
| Q statistic test  SL | 1271.69  0.0000 |
| ζ2 between studies | 0.12 |
| ζ2 within-studies | 0.09 |
| I2 at study level | 53% |
| I2 at effect level | 44% |
| N | 59 |

Figure S2. ‘Multi-level’ model forest plot for T2D

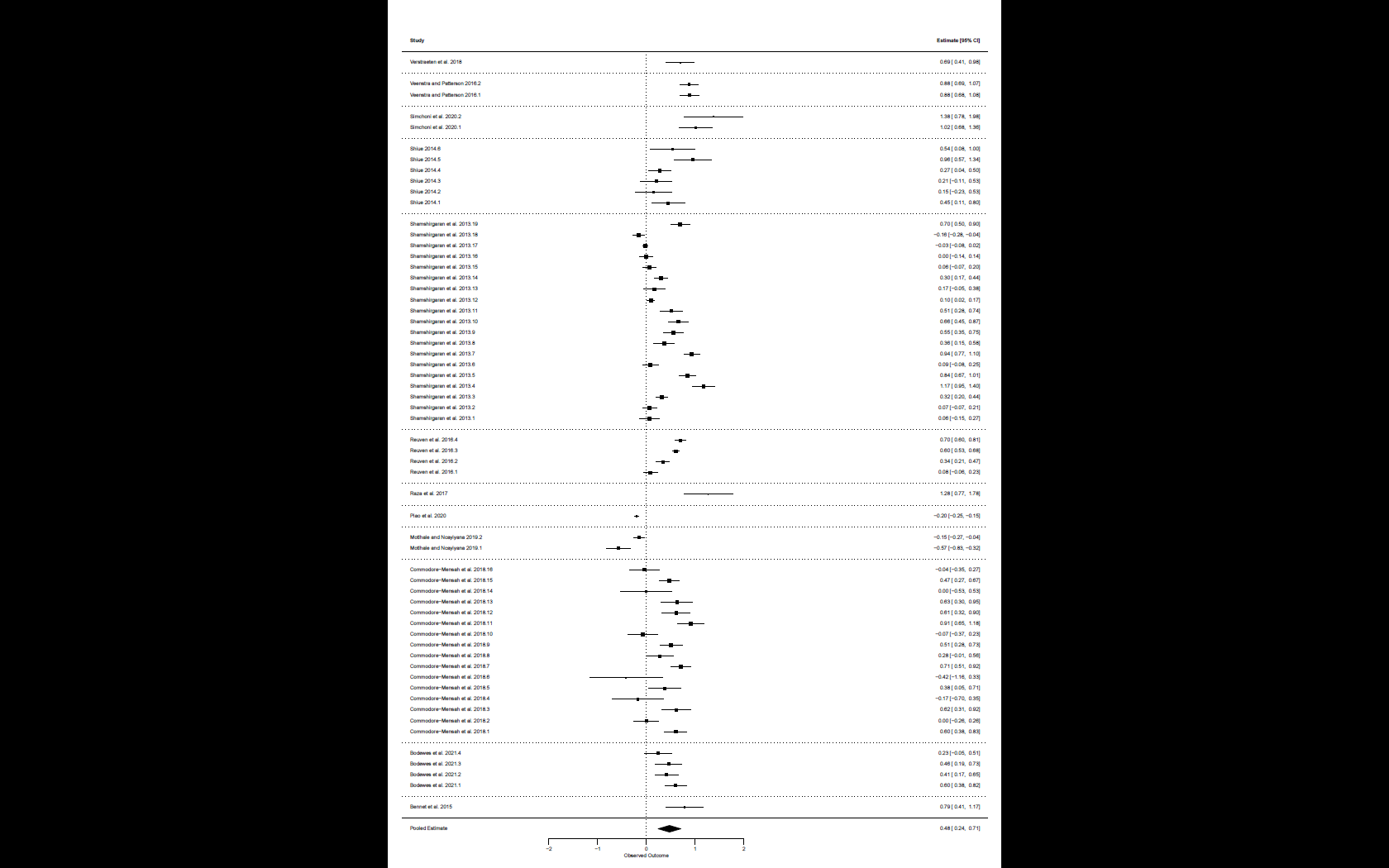


Figure S3: Galbraith plot for T2D, obesity, hypertension, other and child obesity: all sample

T2D Obesity



Hypertension Other



Child obesity



Table S7: Test for bias due to small study effects (Egger’s test)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

T2D Obesity Hypertension Child Obesity

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Panel A: no subgroups

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Beta 1.14 (.814) .45 (.537) -.46 (.948) .98 (.656)

z-value 1.40 .84 -.49 1.49

SigLevel .163 .402 .627 .136

Panel B: Using origin subgroups

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Beta .86 (.812) .40 (.553) -.94 (1.00) .72 (.832)

z-value 1.06 .72 -.94 .87

SigLevel .291 .471 .348 .384

Figure S4a: Estimates of size effects and CI after leave-one-out estimation: T2D



Figure S4b: Estimates of size effects and CI after leave-one-out estimation: Obesity



Figure S4c: Estimates of size effects and CI after leave-one-out estimation: Child obesity



Table S8a: Estimates of effects for T2D before and after performing trim-and-fill estimation

Studies Estimated Effect size 95% confidence interval

Observed 0.398 0.297; 0.498

Observed + Imputed 0.398 0.297; 0.498

N = 59

Observed = 59

Imputed = 0

Figure S5a: Funnel plot of effects for T2D after trim and fill



Table S8b: Estimates of effects for obesity before and after performing trim-and-fill estimation

Studies Estimated Effect size 95% confidence interval

Observed -.112 -0.218; -.006

Observed + Imputed -.242 -0.347; -.138

N = 93

Observed = 75

Imputed = 18

Figure S5b: Funnel plot of effects for obesity after trim and fill



Table S8c: Estimates of effects for hypertension before and after performing trim-and-fill estimation

Studies Estimated Effect size [95% conf. interval]

Observed 0.069 -0.010 0.149

Observed + Imputed 0.069 -0.010 0.149

N=36

Observed=36

Imputed=0

Figure S5c: Funnel plot of effects for hypertension after trim and fill



Table S8d: Estimates of effects for child obesity before and after performing trim-and-fill estimation

Studies Estimated Effect size 95% confidence interval

Observed .224 0.035; .412

Observed + Imputed .145 -0.056; .346

N = 22

Observed = 19

Imputed = 3

Figure S5d: Funnel plot of effects for child obesity after trim and fill



Table S9. Estimates for migrant excess risk using host population as contrast\*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | T2D | Adult obesity | Hypertension | Child Obesity | Other |
| STATISTIC |  |  |  |  |  |
| Theta(logPR) | 0·36 | -0·11 | 0·015 | 0·20 | -0·47 |
| 95% CI | [0·25;0·48] | [-0·21;-0·01] | [-0·06;0·89] | [0·04;0·36] | [-0·74;-0·20] |
| Z  Prob> | 6·27  ·0001 | -2·20  ·0028 | 0·40  ·6881 | 2·45  ·0143 | -3·44  ·0006 |
| I2 | 97·6% | 99·8% | 96% | 40·3% | 91% |
| Q statistic test    SL | 1271  ·0000 | 43509  ·0000 | 1013·44  ·0000 | 36·87  ·0245 | 83·19  ·0000 |
| N | 54 | 76 | 41 | 23 | 10 |

**\*** Excluding studies with OR that could not be converted to PR and including studies that restricted host population to native minority populations.

'Other' includes outcomes not directly invoked in DOHaD as a result of exposures to early conditions: asthma (n=5), cancer (n=3) and mental disorders (n=2).

1. https://github.com/sdaza/zotnote [↑](#footnote-ref-1)