International treaties have mostly failed to produce their intended effects

Steven J. Hoffman1,2,3,4, Prativa Baral1a,6, Sophie Campbell1b,6, Brooke Campust, Maria Dantas3,4, Neda Foroughiana, Gaëlle Groux4, Elliot Gunna, Gordon Guyatti5, Roojin Habibib, Mina Karabita, John N. Lawsc,6,7, Olivia Lee8, Binxi Lia, John-Arne Røttingen9, Nicola Saharon10, Archita Srivastava1b, Ali Tejpard, Maxwell Tran1b, Yu-qing Zhange11, Qi Zhou12, and Mathieu J. Poirier13

There are over 250,000 international treaties that aim to foster global cooperation. But are treaties actually helpful for addressing global challenges? This systematic field-wide evidence synthesis of 224 primary studies and meta-analysis of the higher-quality 82 studies finds treaties have mostly failed to produce their intended effects. The only exceptions are treaties governing international trade and finance, which consistently produced intended effects. We also found evidence that impactful treaties achieve their effects through socialization and normative processes rather than longer-term legal processes and that enforcement mechanisms are the only modifiable treaty design choice with the potential to improve the effectiveness of treaties governing environmental, human rights, humanitarian, maritime, and security policy domains. This evidence synthesis raises doubts about the value of international treaties that neither regulate trade nor contain enforcement mechanisms.

international law | global governance | global legal epidemiology | human rights | systematic review

International treaties are often used by countries to address concerns that transcend national boundaries, including the environment, human rights, humanitarian crises, maritime issues, security, and trade (1–18). Today there are at least 250,000 treaties (19) yet relatively few have been evaluated for impact (20–22), which means we do not know whether these legally binding instruments are effectively serving their intended purpose (23, 24). And yet, leaders from government, academia, business, and civil society routinely call for new treaties to address global challenges (25, 26). Few calls for new treaties consider the costs of drafting, signing, ratifying, and enforcing them. Moreover, the indirect opportunity costs associated with the resources and rhetorical space that go into negotiating and implementing treaties may draw attention away from potentially more important initiatives (23, 25, 26).

Given the central role of treaties in global governance, it is surprising that there has never been a systematic review or field-wide evidence synthesis of whether treaties actually have effects. Individual primary studies have produced divergent results (27–29) which have catalyzed intense debates among researchers and global policymakers on the value of treaties and whether new treaties should be negotiated (28, 30–32). There is even debate on how treaties might theoretically produce effects: for example, through immediate socialization processes during intergovernmental negotiations, short-term normative processes at the time of treaties’ initial adoption, or longer-term legal processes following treaties’ ratification and coming into force (33–35). Very few studies have evaluated the impact of different treaty designs across policy domains, leaving today’s diplomats and political leaders with just personal experience, case studies, and intuition on which to design and negotiate new legal arrangements.

Previous evaluations of treaties found effects to be dependent on the negotiating process, intended outcome, policy domain, and treaty design (20, 23, 30). For example, treaties negotiated through large multilateral organizations composed of over 100 countries like the United Nations (UN) may be less effective than treaties with fewer parties (30). Likewise, changing people’s behavior through treaty making may be more difficult than regulating physical places, government policies, and marketable products (23). Finally, previous evaluations emphasize the potential value of embedding accountability mechanisms within treaties, including transparency mechanisms for information sharing, oversight mechanisms for monitoring parties, complaint mechanisms for adjudicating grievances, and enforcement mechanisms for sanctioning noncompliance (5, 36). Uncertainty around the impacts of these factors points to the need for synthesizing evidence across many primary studies to better understand how treaties can be designed for maximum impact.

Significance

International treaties have mostly failed to produce their intended effects except for international trade and financial laws and treaties with enforcement mechanisms. These results are unexpected because they challenge conventional wisdom about treaties, which are widely considered as the apex mechanism for countries to make commitments to each other. Not only do our findings question the usefulness of the more than 250,000 existing treaties that have been negotiated to date but they should directly inform how national governments and international institutions facilitate global cooperation on the myriad challenges we face and how future international treaties can be better designed for greater impact.

Published August 1, 2022.

Competing interest statement: S.J.H. is the scientific director of the Institute of Population and Public Health at the Canadian Institutes of Health Research. J.-A.R. is the ambassador for global health at the Norwegian Ministry of Foreign Affairs. This article is a PNAS Direct Submission. Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

To whom correspondence may be addressed. Email: steven.hoffman@globalstrategylab.org.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2122854119/DCSupplemental.
We conducted a systematic field-wide evidence synthesis to evaluate the effects of international treaties, which included a rigorous systematic review of all existing quantitative impact evaluations of treaties based on a published protocol (37). Our systematic review searched 24,096 records, of which 224 primary studies met our inclusion criteria and 82 studies reported enough data for meta-analysis and meta-regression analysis. These 82 studies evaluated 53 unique treaties involving more than two countries and included 199 unique quantitative estimates that could be converted into standardized effect sizes to facilitate comparisons across treaties, policy domains, accountability mechanisms, institutional contexts, and study characteristics. Only 2 of the 82 studies were assessed as having serious risk of bias using Cochrane’s ROBINS-I tool for nonrandomized studies (SI Appendix, Fig. S2), with bias due to confounding and missing data most commonly affecting assessments. As part of our analysis, we distinguished studies that found effects in the treaty’s intended direction as compared with studies that found effects counter to the intended direction. The 142 studies that provided insufficient data for further statistical analysis or examined bilateral treaties were evaluated using an “extreme case” mixed-methods approach with sequential design and parallel sampling strategy (38, 39). Integrating these techniques (40) allows us to draw inferences from all 224 studies that together represent the most comprehensive collection of evidence yet compiled on treaties’ effects.

Here, we use these techniques to evaluate studies of which treaties have effects, what those effects are, and how future treaties could be designed for greater effectiveness. Applying a scientific approach to the centuries-old field of international law has the potential to change the way we use and study the treaties that govern nearly every aspect of our lives and on which addressing innumerable global challenges depends.

Results

We found considerable variation in treaty effects (Box 1). This heterogeneity means that a single synthesized effect of all treaties is not informative; rather, to better understand which treaties work and which features are associated with greater effectiveness, treaties need to be grouped by characteristics and evaluated accordingly, as we have done with our meta-analyses (Fig. 1).

Impacts by Policy Domain. We measured overall effects of treaties in six policy domains; of these domains, only trade and finance treaties were associated with statistically significant intended effects (Z = 3.60; CI 2.42 to 4.78). In contrast, effect sizes for treaties governing the environment, human rights, humanitarian crises, maritime issues, and security were not statistically significant (Table 1). As an alternative approach, meta-analysis of the types of outcomes being evaluated by the studies (as opposed to the policy domains in which the treaties were negotiated) produced nearly identical results; only evaluations of international treaties’ impacts on economic outcomes produced statistically significant effects in the intended direction (Z = 3.33; CI 2.10 to 4.56), while studies evaluating treaties’ civil liberties and social outcomes did not generate significant effects. Meta-analysis by the intended locus of change supports these findings, with statistically significant effects identified for changes in products (Z = 3.84; CI 2.52 to 5.16) but not changes in people, places, or policies.

Examining the distribution of standardized effect sizes for each treaty studied further reinforces these consistent findings (Fig. 2 and SI Appendix, Table 8). Even among social and civil liberties outcomes, it was the General Agreement on Tariffs and Trade that was associated with two of the largest intended impacts, namely reducing rates of child labor and easing migration to new member countries (41, 42). These cross-sectoral

---

Box 1. Summary of the 224 included studies by outcomes evaluated

| Category             | Summary of the 224 included studies                                                                                                                                                                                                 |
|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Overall              | A systematic search of 11 databases identified 24,096 records, of which 224 studies were quantitative impact evaluations of treaties (SI Appendix, Fig. S1); 82 studies including 199 unique quantitative estimates of effects were meta-analyzed. Qualitative data were extracted from the remaining 142 studies that provided insufficient data for further statistical analysis or examined bilateral treaties. On average, 29.6 y passed between a treaty's adoption and the publication of an evaluation (range 3 to 67 y). Economic outcomes were the most frequently evaluated (n = 82/199). |
| Civil liberty outcomes | The General Agreement on Tariffs and Trade (Z = 5.98) and the Optional Protocol to the International Covenant on Civil and Political Rights (Z = 1.62) produced the largest intended effects on civil liberties, while the UN Convention on the Rights of the Child (Z = -2.77) and the International Covenant on Economic, Social and Cultural Rights (Z = -1.44) produced the largest effects opposite the treaty's intentions. |
| Economic outcomes    | The treaties creating the Organization for Economic Co-Operation and Development (Z = 22.75) and Association of Southeast Asian Nations (Z = 14.06) generated the largest intended effects on economic outcomes, while the Common Market for Eastern and Southern Africa (Z = -0.83) and the South Asian Association for Regional Cooperation Preferential Trading Agreement (Z = -0.50) were found to produce the smallest intended effects. |
| Social outcomes      | The Convention on International Trade in Endangered Species (Z = 8.57) and General Agreement on Tariffs and Trade (Z = 10.81) produced the largest intended effects on social outcomes, while the Optional Protocol to the International Covenant on Civil and Political Rights (Z = -3.29) and the Inter-American Convention to Prevent and Punish Torture (Z = -3.26) produced effects opposite the treaty's intended purpose. |

Overview of key empirical findings featuring the largest average treaty effects identified for each type of outcome.
effects were not found among treaties that do not primarily govern trade and finance, with International Labor Organization membership having no effect on unemployment insurance spending (43) and the Convention on the Elimination of All Forms of Discrimination against Women having no effect on women’s economic rights (44).

Studies that met our inclusion criteria but lacked sufficient data for meta-analysis further support the impact of treaties on economic outcomes, particularly international trade. For example, several studies found the North American Free Trade Agreement increased foreign direct investment (45, 46) and intraregional trade (47); World Trade Organization agreements increased commitments to liberalizing tariffs and improving market access, particularly among developing countries (48); and economic integration agreements (49) and foreign trade agreements (50) promoted trade flows. Similarly, the Treaty of Asunción that created the Mercosur common market has been linked to increased trade flows among participating South American countries (51, 52) and with China specifically (53), increasing bloc competitiveness with nonmember countries (54), and greater commitment from member countries to the observance of human rights (52, 55, 56).

Unintended and Harmful Treaty Impacts. In contrast to considerable evidence in support of trade and finance treaties’ effectiveness, human rights and environmental agreements have not been shown to consistently improve state practices (57) and, as many studies conclude (27, 29, 32, 44, 58–63), may actually have led to unintended harmful effects such as increased torture (64, 65). As one striking example of potential unintended consequences, the UN Convention on the Rights of the Child was found to be the treaty with the most harmful effects \( (Z = -2.77; CI = -15.38 to 9.84) \) among the 32 treaties that were evaluated by at least two studies (Fig. 2). Ratification of this treaty has been associated with lower Amnesty International human rights ratings (66), no improvements in health outcomes (67), worsened human rights practices (28), and, paradoxically, increases in child labor (58).

Findings from studies that were not able to be included in the meta-analysis point to similarly neutral or negative effects, in which the ratification of human rights treaties showed no relation to improvements in health or social outcomes (67, 68), in part because treaty ratification does not necessarily lead to better human rights practices (44, 66). One hypothesis that has been proposed for this “paradox of empty promises” is that repressive governments face few negative consequences and reap diplomatic rewards for signing human rights treaties without meaningfully implementing provisions that may be counter to their interests (20, 28).

Designing More Effective Treaties. We found evidence that including enforcement mechanisms within treaties led to increased effectiveness \( (Z = 3.13; CI 1.65 to 4.61) \), but that transparency, complaint, and oversight mechanisms did not (Fig. 2). Importantly, enforcement mechanisms were included in 9 of the 20 trade and finance treaties evaluated but in only two of the five environmental treaties and none of the 28 evaluated treaties governing human rights, humanitarian crises, maritime issues, and security (SI Appendix, Table S2). Nevertheless, trade and finance treaties with enforcement mechanisms were not associated with greater effectiveness than trade and finance treaties without enforcement mechanisms, while the two environmental treaties with enforcement mechanisms were found to be more effective than other treaties (SI Appendix, Fig. S4 and Table S4). These findings suggest the impact of enforcement mechanisms might be contingent on a treaty’s policy domain and might be especially helpful for treaties outside of trade and finance.

Turning to negotiating forums, only treaties negotiated through economic cooperation organizations like the Organization for Economic Co-Operation and Development or World Trade Organization were found to produce statistically significant effects \( (Z = 3.19; CI 2.06 to 4.32) \), while those negotiated through UN agencies and human rights organizations did not (Fig. 2). The size of a treaty’s negotiating forum also appears to

### Fig. 1. Forest plots of standardized effect sizes by variable groupings. Forest plots detailing (A) effects by policy domain; (B) accountability effects for transparency, oversight, complaint, and enforcement mechanisms; (C) institutional effects including the year of treaty adoption, type of negotiating forum, and size of the negotiating forum; and (D) study effects, including the outcomes evaluated, locus of evaluated change, and when the treaty was evaluated. The width of each bar indicates the 95% CI. The number above each bar indicates the number of study estimates included in the analysis. CIs that do not contain zero indicate a statistically significant effect at the 95% confidence level.
influence outcomes, with a suggestive dose–response relationship between increasing forum size and reduced impacts in the intended direction (Table 1 and SI Appendix, Fig. S3). This relationship might stem from a gradual weakening of treaty provisions during consensus-based negotiations when more countries are involved and there are more divergent positions that need to be accommodated (69). Alternatively, this finding might relate to the type of trade and finance treaties that are typically negotiated through economic cooperation organizations and the smaller size of these forums (SI Appendix, Table S5).

When Treaties Have Impacts. When grouped together, only treaties adopted post-1990 were found to achieve statistically significant intended effects ($Z = 2.73; CI 1.44 to 4.03$), unlike treaties adopted pre-1970 ($Z = 1.18; CI −0.23 to 2.60$) and from 1970 to 1989 ($Z = 0.85; CI −0.64 to 2.33$), but these between-group differences were not statistically significant in the meta-analysis. This difference may be explained by the rising proportion of meta-analyzed treaties governing trade and finance after 1990 (19.7% pre-1990; 80.5% post-1990), such as the General Agreement on Tariffs and Trade and the North American Free Trade Agreement. However, effect sizes did vary according to the point at which treaties were evaluated by researchers. Studies that evaluated the immediate socialization effects of a treaty’s negotiation or the short-term normative effect of a treaty’s adoption found the largest effects ($Z = 3.78; CI 2.02 to 5.53$), while studies that evaluated the longer-term legal effects of a treaty’s ratification ($Z = 1.20; CI 0.25 to 2.14$) and its coming into force ($Z = 0.67; CI −2.14 to 3.49$) found considerably smaller effects and no effects, respectively. This finding provides evidence that impactful treaties achieve their effects through immediate socialization and short-term normative processes rather than longer-term legal processes.

As one example, studies evaluating the UN Convention against Torture (CAT) found different levels of effectiveness depending on whether socialization, normative, or legal effects were being evaluated: The six most positive findings came from evaluations at the earliest stage of the treaty’s adoption, whereas

| Table 1. Meta-analysis results |
|--------------------------------|
| **Group** | **Meta-analyses** | **Variables** | **Treaties, n** | **Estimates, n** | **Z statistic** | **95% CI** | **P value** |
|---------|-----------------|---------------|----------------|-----------------|----------------|------------|------------|
| Policy domains | 1. Environmental treaties | Yes | 5 | 13 | 1.64 | (−1.52, 4.81) | 0.978 |
| | 2. Human rights treaties | Yes | 48 | 186 | 1.69*** | (0.85, 2.53) | 0.001*** |
| | 3. Humanitarian treaties | Yes | 14 | 80 | 0.05 | (−1.19, 1.30) | 0.64 |
| | 4. Maritime treaties | Yes | 39 | 119 | 2.78*** | (1.26, 3.80) | 0.029 |
| | 5. Security treaties | Yes | 4 | 8 | −0.41 | (−4.44, 3.61) | 0.297 |
| | 6. Trade and finance treaties | Yes | 6 | 8 | −0.04 | (−4.07, 3.99) | 0.392 |
| Accountability | 7. Transparency mechanisms | Present | 43 | 182 | 1.78*** | (0.93, 2.63) | 0.479 |
| | 8. Oversight mechanisms | Present | 10 | 17 | 0.76 | (−1.93, 3.45) | 0.597 |
| | 9. Complaint mechanisms | Present | 12 | 33 | 1.74* | (−0.35, 3.82) | 0.188 |
| | 10. Enforcement mechanisms | Present | 27 | 115 | 2.15*** | (1.09, 3.21) | 0.023* |
| Institutional effects | 11. Size of negotiating forum | Large ($n \geq 100$) | 25 | 136 | 1.12** | (0.16, 2.08) | 0.049* |
| | | Medium ($5 < n < 100$) | 24 | 42 | 2.26** | (0.52, 4.00) | 0.001*** |
| | | Small ($n \leq 5$) | 4 | 21 | 4.36*** | (1.84, 6.88) | 0.001*** |
| | 12. Type of negotiating forum | Economic cooperation | 27 | 96 | 3.19*** | (2.06, 4.32) | 0.001*** |
| | | Human rights | 5 | 9 | −0.86 | (−4.55, 2.83) | 0.120 |
| | | UN agency | 21 | 94 | 0.39 | (−0.75, 1.54) | 0.210 |
| | 13. Year of treaty adoption | Pre-1970 | 19 | 64 | −1.18 | (−2.30, 0.26) | 0.049* |
| | | 1970 to 1989 | 18 | 37 | 0.85 | (−0.64, 2.33) | 0.001*** |
| | | Post-1990 | 17 | 77 | 2.73*** | (1.44, 4.03) | 0.001*** |
| Study effects | 14. Locus of evaluated change | Change in people | 12 | 23 | 0.94 | (−1.36, 3.24) | 0.001*** |
| | | Change in places | 5 | 12 | 0.63 | (−2.56, 3.81) | 0.64 |
| | | Change in policies | 24 | 94 | 0.40 | (−0.74, 1.53) | 0.001*** |
| | 15. Outcomes evaluated | Civil liberty outcomes | 12 | 44 | 0.46 | (−1.22, 2.13) | 0.003*** |
| | | Economic outcomes | 22 | 82 | 3.33*** | (2.10, 4.56) | 0.001*** |
| | | Social outcomes | 31 | 73 | 0.58 | (−0.72, 1.88) | 0.001*** |
| | 16. When treaty was evaluated | Ratification | 43 | 142 | 1.20** | (0.25, 2.14) | 0.001*** |
| | | Coming into legal force | 11 | 16 | 0.67 | (−2.14, 3.49) | 0.001*** |

Sixteen meta-analyses of standardized $Z$ statistics, 95% CI, and $P$ values for each $Z$ statistic and overall effect are presented. Each variable subgroup represents an independent meta-analysis of that subgroup’s effect size. Cis that do not contain zero indicate a statistically significant effect, while a $P$ value below 0.05 indicates a statistically significant difference between meta-analyzed variables, at the 95% confidence level. **$P < 0.01$, ***$P < 0.05$, *$P < 0.1$. 
the most negative study findings evaluated the CAT at the later stage of ratification (SI Appendix, Table S8). For example, a study evaluating the CAT’s impact 1 year after the treaty’s signing found that signatories were less likely to torture and that citizens of those countries experienced fewer civil war deaths (70). In contrast, the study finding the largest negative effect matched CAT-ratifying countries with nonratifying countries, finding that the treaty increased the probability that ratifying countries would practice torture by 15% (44). These seemingly contradictory results may indicate that authoritarian governments are likely to sign treaties like the CAT for immediate political advantages but, in the absence of strong enforcement mechanisms, early intended impacts of treaty signing can be reversed by the time of treaty ratification (28).

**Determinants of Treaty Effectiveness.** Two key determinants of effectiveness were identified through a meta-regression analysis of all 199 standardized effect estimates from the 82 higher-quality studies of 53 treaties on all variables for policy domains, accountability mechanisms, institutional contexts, and study characteristics: 1) a treaty governing trade and finance, and 2) a study evaluating a treaty’s short-term normative effects (Table 2). Alternative models evaluating economic outcomes (β = 2.22; CI 0.17 to 4.28), economic cooperation negotiating venue (β = 2.43; CI 0.40 to 4.47), or a change in products (β = 2.91; CI 0.95 to 4.87) produced similar but slightly weaker effects (SI Appendix; Table S6) than our primary model that privileged a treaty’s policy domain (trade and finance: β = 3.23; CI 1.11 to 5.36). The increased impact seen when evaluating a treaty’s normative effects is statistically significant in most models except when paired with a variable for trade and finance treaties. Our meta-regression results on enforcement mechanisms diverge from subgroup meta-analyses, suggesting these accountability mechanisms are not independent drivers of treaty effectiveness (β = 0.04; CI -2.19 to 2.28).

**Discussion**

**Key Findings.** This systematic field-wide evidence synthesis finds that international treaties have mostly failed to produce their intended effects. Trade and finance is the only policy domain with considerable and consistent evidence of intended impacts. The numerous treaties governing other policy domains have either not worked or have been insufficiently studied with rigorous methods to demonstrate positive impacts. These divergent results are reinforced by similarly strong evidence of treaties producing intended impacts when evaluating economic outcomes, changes in products, and when treaties were negotiated through economic cooperation forums. Although these results may be partly attributable to more easily measured quantitative outcomes, trade and finance treaties seem to be consistently effective no matter how they are designed or evaluated.

For treaties governing environmental, human rights, humanitarian, maritime, and security policy domains, the only modifiable treaty design choice with the potential to improve effectiveness appears to be the inclusion of enforcement mechanisms. The importance of enforcement mechanisms such as prescribing financial sanctions on countries or expelling countries from treaty bodies and trade blocs is supported by research on compliance with international law (71). However, even well-designed accountability mechanisms may not be enough to overcome...
compliance challenges in some policy domains (32). In contrast, it appears that complaint, oversight, and transparency mechanisms are not associated with greater treaty effectiveness. Yet only 2 of the 32 treaties that were not governing trade and finance made use of enforcement mechanisms, meaning that this accountability mechanism is likely underutilized and that its role in determining treaties’ effects should be further investigated. Smaller negotiating forums may also drive intended effects, but it is difficult to statistically disentangle the effects of a forum’s size from the influence of its policy domain.

Finally, the point at which treaties were evaluated by researchers was one of the few statistically significant variables used in our meta-analyses. The larger intended effects observed when evaluating treaties at the time of their signing suggests that immediate socialization and short-term normative processes stemming from treaties’ negotiation may be more important than the longer-term legal processes stemming from their ratification and coming into legal force. Alternatively, treaty effects might diminish or become less measurable over time, might appear larger when based on short-term technical outputs rather than socially relevant outcomes, or might be driven by early-adopting countries being more willing to comply with international law.

While the 224 studies analyzed in this evidence synthesis constitute a substantial body of scientific literature on the impact of international treaties, the quality and breadth of this evidence must improve. More research should focus on disentangling the contexts and circumstances in which treaty design choices like accountability mechanisms can be effectively deployed to maximize intended impacts. If the power of treaties’ negotiating process is confirmed, then additional research on how to effectively design, navigate, and leverage this political process for catalyzing change would be very productive.

**Conclusion**

Unless different evidence emerges, calls for new international treaties to address global challenges beyond trade and finance should be received with caution. Although the meta-analysis relies on the current state of published evidence, our findings that treaties governing environmental, human rights, humanitarian, maritime, and security policy domains have not demonstrated impacts either point to the failure of these treaties to achieve impacts or the failure of researchers to generate evidence of impacts. If pursued, enforcement mechanisms appear to be the only treaty design choice that holds promise of maximizing the chances of achieving intended effects. Future treaties beyond trade and finance that do not have enforcement mechanisms are unlikely to be worth their considerable effort and may have unintended consequences. These findings are immediately relevant for treaties that are currently being negotiated or that are being considered for negotiation.

**Methods and Materials**

The protocol detailing the methods and design of this systematic field-wide evidence synthesis of effects is registered with the PROSPERO international prospective register of systematic reviews (registration number CRD42015019830) and published in Systematic Reviews (37). In summary, all quantitative impact evaluations that measured the effects of an international treaty involving more than two countries on an objective and quantifiable outcome were included. Study “participants” included all current and former countries.

Ten electronic bibliographic databases were searched from inception to December 2017: Applied Social Sciences Index & Abstracts, CINAHL, Global Health, International Bibliography of Social Sciences, International Political Science Abstracts, MEDLINE, PAIS International, Social Sciences Abstracts, Social Sciences Citation Index, and Worldwide Political Science Abstracts. The exact search strategies implemented for each database are presented in SI Appendix, Dataset S1 and a description of search results is presented in Box 1. The study selection included publicly available peer-reviewed studies and gray literature (e.g., dissertations) that aimed to quantitatively measure an international treaty’s effect on any objective and quantifiable outcome (24). Further details on the review of titles and abstracts for eligibility, final consensus for inclusion, aggregation and deduplication using EndNote, and assessment for bias using Cochrane’s ROBINS-I tool are found in SI Appendix.

**Meta-Analysis.** Hoffman and Ratttingen’s earlier scoping review of 90 quantitative impact evaluations (23) included studies that were mostly time-series cross-sectional analyses with quantitative data on continuous variables suitable for pooling. However, these studies evaluated treaties across different countries, years, and outcome variables. Indeed, most study data of effect sizes were measured in different scales due to this diversity of outcomes, treaties, study designs, and study populations.

To address this problem, we deviated from the published protocol and transformed all outcomes to standardized $T$ statistics, which offer the advantage of being robust to heteroscedasticity of effect sizes across studies due to different sample sizes and precision (72). Standardized effect sizes are functions of estimates of the regression coefficient ($β$), odds ratio (OR), or other effect measure, and of $SEs (σ)$, $P$ values, or other measures of variance. These $T$ statistics were further approximated to standardized normal $Z$ statistics assuming the sample size of each analysis equaled or exceeded 30. Each $Z$ statistic was assigned a positive value if the direction of effect was found to go in the intended direction for

### Table 2. Meta-regression results by variable subgroup

|                  | Model 1       | Model 2       | Model 3       | Model 4       |
|------------------|---------------|---------------|---------------|---------------|
| Adopted post-1990 | 1.71**        | 0.89          | 0.82          | −0.45         |
|                  | (0.06, 3.37)  | (−1.18, 2.96) | (−1.23, 2.87) | (−2.63, 1.72) |
| Enforcement mechanism | 1.47          | 1.22          | 0.08          |               |
|                  | (−0.75, 3.69) | (−0.98, 3.43) | (−2.21, 2.37) |               |
| Study evaluated the treaty’s signing | 2.34**        |               | 1.65          |               |
|                  | (0.35, 4.33)  |               | (−0.35, 3.65) |               |
| Trade and finance treaty |               |               | 3.25***       | (1.10, 5.40)  |
| Constant         | 1.02*         | 0.91*         | 0.53          | 0.09          |
|                  | (−0.01, 2.05) | (−0.13, 1.96) | (−0.55, 1.61) | (−1.00, 1.19) |
| Observations     | 199           | 199           | 199           | 199           |
| Adjusted $R$-squared | 0.016         | 0.020         | 0.042         | 0.081         |

Random-effects meta-regression results using aggregate-level data examining significant predictors of standardized effect sizes across variable subgroups. Variable inclusion for the full model was determined iteratively by the highest adjusted $R$-squared value, with positive values indicating stronger treaty impacts in the intended direction. Regression coefficients are reported with 95% CI in parentheses. *** $P < 0.01$, ** $P < 0.05$, * $P < 0.1$. 
each treaty, and a negative value if the direction of effect was found to point in the opposite direction from what was intended for each treaty; a value of zero would indicate no difference from the study mean. As a measure of the number of SDs above or below the study mean, these Z statistics can be compared as standardized values. A detailed summary of studies included in the analysis can be found in SI Appendix, Table S7.

**Subgroup Analysis.** We further analyzed our extracted data through subgroup analysis using χ² tests of heterogeneity. Specifically, informed by an analytic framework we previously developed (23, 26), we examined the following potential determinants of heterogeneity, including postulated directions of effect. Based on Hoffman and Rettingen's recent assessment of international treaties' effects (23), we conducted nine primary categories of subgroup analyses, some of which deviated from the previously published protocol, to ascertain these factors' impacts on the effectiveness of international treaties:

1) By study design and effect relative to purpose of treaty: The study designs included in this review included cross-sectional studies, time-series studies of individual jurisdictions, and time-series studies of multiple jurisdictions. Subgroup analyses of impacts in either the intended or opposite direction for each treaty were analyzed to determine whether there were statistically significant effects for the direction, or if the meta-analysis resulted in no statistically significant effects at all. Our findings point to competing significant effects in each of these directions rather than a lack of statistically significant effects in either the positive or negative direction and no significant difference between study designs (SI Appendix, Table S3). This finding is further supported by the studies that met our inclusion criteria but were not able to be meta-analyzed.

2) By time period: An international treaty's time period of initial adoption may give insights into whether enough time has elapsed to determine its impacts as well as whether international treaties adopted in different eras might have had stronger impacts. Studies were divided according to whether the international treaties under study were ratified prior to 1970, from 1970 to 1989, or after 1990. We found no statistically significant difference on strength of impacts over time, despite a generally increasing trend in intended treaty impacts for both the year the study was conducted and the year the treaty was signed (SI Appendix, Fig. S3).

3) By when the treaty was evaluated: The stage at which a treaty is evaluated can have an impact on reported effects. Studies were coded to have evaluated the effects of a treaty at either the stage of a) signing, b) ratification, or c) coming into legal force. Both immediate socialization processes during intergovernmental negotiations and short-term normative processes at the time of treaties' initial adoption (as opposed to long-term changes in norms and discourse) may contribute to effects observed at the time of treaty signing. Longer-term legal processes may also contribute to effects observed following the ratification or coming into force of a treaty within countries.

4) By size of the negotiating forum: There may be a relationship between the number of countries that are party to a treaty and the impact measured on outcomes. This criterion allowed for comparisons between international treaties binding a small group of countries versus treaties with many parties (SI Appendix, Fig. S3). Subgroups were created for small (n ≤ 5), medium (5 < n < 100), and large (n ≥ 100) treaties.

5) By type of negotiating forum: The forum through which a treaty was negotiated may provide information on whether treaties have differing effects depending on the process and/or hosting organization. Studies were therefore grouped according to whether they evaluated a treaty that had been negotiated through an economic cooperation forum, non-UN humanitarian forum, or UN forum.

6) By accountability mechanism: Level or robustness of accountability mechanisms embedded in treaties or used by states to encourage compliance such as legal penalties, monetary fines, and “naming and shaming” may impact the magnitude of its effects. To capture these effects, each international treaty under study was coded for whether it contained any textual provisions that created accountability mechanisms for transparency, complaint, oversight, or enforcement (5). Transparency mechanisms enable information to be shared about countries with observers through regular reporting and information aggregation. Oversight mechanisms build on transparency by actively monitoring and evaluating countries through standard setting and implementation review. Complaint mechanisms allow grievances attributable to a country to be processed and adjudicated through country or secretariat complaint mechanisms. Enforcement mechanisms foresee the possibility of a specific sanction or consequence delivered by a court, committee, secretariat, or other legal authority, even if that authority was not created by the treaty.

7) By locus of evaluated change: Outcomes were broadly categorized as changes to people (e.g., health status), places (e.g., carbon emissions), policies (e.g., enacted regulations), or products (e.g., trade flows) (SI Appendix, Table S7). This category helped evaluate whether, for example, changing people’s behavior through treaty making may be less effective than regulating products or places.

8) By outcomes: Analysis of the types of outcomes evaluated by studies allowed for the comparison of similar types of impacts, regardless of a given treaty's intended primary outcomes. Study outcomes were aggregated at broad levels of civil liberties, social outcomes, and economic outcomes.

9) By policy domain: Treaties were separately classified into one of six categories of core treaty objectives: environmental, human rights, humanitarian, maritime, security, and trade and finance. The policy domains in which treaties were negotiated were coded separately from the outcomes evaluated by studies meeting inclusion criteria.

Using code provided in SI Appendix, Dataset S2, meta-analysis pooled Z statistics using each of these categories and CI were constructed using the HK3 method (73). Meta-analysis demonstrated a statistically significant effect of an international treaty if the derived 95% CI excludes zero. We examined the extent of heterogeneity statistically and of inconsistency using the I squared (I²) measure.

Meta-analysis of all 199 unique quantitative estimates that could be converted into standardized effect sizes evaluating 53 unique treaties involving more than two countries resulted in a Z statistic of 1.69 (CI 1.55 to 1.82) with an I² of 97.0% indicating considerable heterogeneity. This aggregate effect was not reported as a primary study outcome because the statistically significant positive effect size is entirely driven by the strong influence of trade and finance treaties and not by the null effects of the five other policy domains.

**Meta-Regression Analysis.** Standardized effect sizes were further analyzed using aggregate data random-effects meta-regression analysis with Knapp-Hartung variance modification using the metareg Stata command (74, 75). Departures from methods outlined in the protocol are due to the use of Z statistics rather than standardized mean difference as the measure of effect size. Because of this change, there were no sample sizes associated with the studies, and no study weighting could be applied to the meta-regression. In addition to regressing by subgroup variables, different variable subgroups were combined to produce the model with the highest possible adjusted R-squared value. Further robustness checks testing for the presence of interaction effects failed to demonstrate statistically significant synergistic effects between trade and finance treaties, changes in products, or economic outcomes. Full results are reported in Table 2 and SI Appendix, Table S1.

**Data from Studies Not Meta-Analyzed.** In order to triangulate our quantitative findings, qualitative descriptions of studies that met the inclusion criteria but 1) evaluated bilateral treaties only, 2) demonstrated aggregated effects of multiple treaties together, 3) lacked the required quantitative information to conduct a meta-analysis, or 4) were conducted as part of a graduate student dissertation were nonetheless incorporated into our findings even if they did not contribute to the meta-analysis. After conducting quantitative analysis, we examined only those international treaties generating the most significantly positive and significantly negative treaty effects and synthesized qualitative data from the entire population of studies (including those meeting the threshold for meta-analysis) for inclusion into the manuscript.

We therefore used an extreme case mixed-methods approach with a sequential design and parallel sampling strategy (38, 39). The extreme case approach was chosen to explore whether qualitative descriptions of studies meeting inclusion criteria but lacking one of the four additional criteria for meta-analysis outlined above would support or contradict quantitative findings for treaties that demonstrated the most significant effects in both the intended and opposite directions for each meta-analysis subgroup. The flow from the dominant quantitative analysis to exploratory qualitative analysis allows us to provide more depth of understanding of the extreme cases identified in the meta-analysis and potentially to introduce contradictory qualitative evidence. Quantitative and qualitative
analyses drew from the same population but were conducted in parallel by different research team members to reduce the risk of bias. In addition to the qualitative evidence included in the manuscript, a detailed summary of all qualitative descriptions can be found in SI Appendix, Table S8.

Strengths and Limitations. This evidence synthesis used a rigorous systematic methodology to investigate an important question that has not yet benefited from systematic review and meta-analysis techniques, informing a particularly salient and active discussion regarding the development of new international treaties (69, 76). Another notable strength of the evidence synthesis is its inclusive and comprehensive scope, including gray literature such as dissertations and theses. Recognizing that many studies, particularly the gray literature, were not able to be included in the meta-analysis, we triangulated the results from these studies into the main analysis.

An important limitation of this evidence synthesis is the breadth of treaty outcomes. Our review included many studies evaluating a wide range of factors influenced by international treaties. Our application of meta-analysis and metaregression techniques to these diverse studies and outcomes may have been affected by the significant heterogeneity among the pool of eligible studies. This heterogeneity posed challenges in comparing the success of different international treaties due to their varying policy domains and characteristics; however, our calculation of effect sizes using standardized Z statistics allowed us to pool results across different outcomes and units of measurement. Some studies may also have been underpowered, adding to the limitations.

Evidence syntheses are always limited by the quantity and quality of the primary studies they integrate, and the choices made by the researchers who led the included studies. For example, it is possible that the kinds of economic outcomes that might be expected from trade and finance treaties are more easily measured, produce more easily quantifiable effects, or are more consistently studied using high-quality quantitative methods. Our search protocols may also have missed some relevant studies not indexed in the bibliographic databases searched or that were not retrieved with the search terms used.

Finally, we caution against strictly causal interpretations of intended and unintended treaty impacts, as study designs included in the meta-analysis vary in their ability to establish causal identification and may be affected by selection issues (77). Despite these potential limitations, this evidence synthesis represents the best available systematic evaluation of the universe of published studies evaluating the effectiveness of international treaties.

Data Availability. All study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. We thank A. Agarwal, N. Jedrezik, V. Lui, N. Natt, J. Petropoulos, and J. Syrotuk for their help with developing the search strategy; P. Alexander for contributing to risk of bias assessment; as well as M. Chen (Chinese), A. Yu (Chinese), W. Boer (Dutch), T. Huijts (Dutch), J. Bailey (German), S. Schandelmaier (German), C. Barbui (Italian), A. Iorio (Italian), A. Skordai (Greek), and J. J. Yepes-Nuñez (Spanish) for their assistance with text translations.

Funding was provided by Research Council of Norway’s Global Health & Vaccination Programme GLOBALVAC Project 234608 (to S.J.H.); Canadian Institutes of Health Research 172982 (to S.J.H.); and Ontario Government’s Ministry of Research, Innovation & Science ER16-12-197 (to S.J.H.). The views expressed in this article are those of the authors and do not necessarily reflect those of their institutions or governments.

Author affiliations: *Global Strategy Lab, Dahdaleh Institute for Global Health Research, School of Global Health, York University, Toronto, ON, M3J 1P3, Canada;* Osogoode Hall Law School, Toronto, ON, M3J 1P3, Canada; *McMaster Health Forum, McMaster University, Hamilton, ON, L8S 4L8, Canada; *Department of Health Research Methods, Evidence, and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, L8S 4L8, Canada; *Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 21205, *AfCenta for Europe, University of Johannesburg, Johannesburg, 2002, South Africa; *Norwegian Ministry of Foreign Affairs, Oslo, 0114, Norway; and *Ningho Nottingham GRADE (Grading of Recommendations, Assessment, Development, and Evaluation) Center, University of Nottingham, Ningbo, 315104, China.

Author contributions: S.J.H. designed research; S.J.H., P.B., S.R.V.K., L.S., M.H., R.G.L., N.F., G. Groux, E.G., G. Guyatt, R.H., M.K., A.K., K.N., J.-A.R., A.T., M.T., N.S., A.S., Y.-q.Z., Q.Z., and M.J.P.P. performed research; S.J.H., S.R.V.K., G.L., G.G., Q.Z., and M.J.P.P. contributed new reagents/analytic tools; S.J.H., G.L., N.F., and M.J.P.P. analyzed data; and S.J.H., P.B., S.R.V.K., L.S., M.H., R.G.L., N.F., G. Groux, E.G., G. Guyatt, R.H., M.K., A.K., K.N., J.-A.R., A.T., M.T., N.S., A.S., Y.-q.Z., Q.Z., and M.J.P.P. wrote the paper.

1. S. Dinhar. Health policy: Regulate alcohol for global health. Nature 482, 302 (2012).
2. S. J. Hoffman, An international legal framework to address antimicrobial resistance. Bull. World Health Organ. 93, 66 (2015).
3. S. J. Hoffman et al., Strategies for achieving global collective action on antimicrobial resistance. Bull. World Health Organ. 93, 867–876 (2015).
4. S. Andersen, S. J. Hoffman, Much can be learned about addressing antibiotic resistance from multilateral environmental agreements. J. Law Med. Ethics 43 (suppl. 3), 44–52 (2015).
5. S. J. Hoffman, T. Ottersen, Addressing antibiotic resistance requires robust international accountability mechanisms. J. Law Med. Ethics 43 (suppl. 3), 53–64 (2015).
6. J. Hoffman, J.-A. Røttingen, J. Henk, International law has a role to play in addressing antibiotic resistance. J. Law Med. Ethics 43 (suppl. 3), 65–67 (2015).
7. A. Behdman, S. J. Hoffman, M. Pearcey. Some global policies for antibiotic resistance depend on legally binding and enforceable commitments. J. Law Med. Ethics 43 (suppl. 3), 68–73 (2015).
8. Z. Razvi, S. J. Hoffman. Effective global action on antibiotic resistance requires careful consideration of convening forums. J. Law Med. Ethics 43 (suppl. 3), 74–78 (2015).
9. S. J. Hoffman, K. Outterson. What will it take to address the global threat of antibiotic resistance? J. Law Med. Ethics 43, 363–368 (2015).
10. C. Árdal et al., International cooperation to improve access to and sustain effectiveness of antimicrobials. Lancet 387, 296–307 (2016).
11. L. O. Gostin, Non-communicable diseases. Healthy living needs global governance. Nature 511, 147–149 (2014).
12. The Lancet, Fighting fake drugs. The role of WHO and pharma. Lancet 377, 1626 (2011).
13. C. Kohler, A. Makady, Harnessing global health diplomacy to curb corruption in health. J. Health Dipl. 1, 1–14 (2013).
14. A. D. Ozman et al., A framework for mandatory impact evaluation to ensure well-informed public policy decisions. Lancet 375, 427–431 (2010).
15. The Lancet, Unequal access: A framework convention for obesity control. Lancet 378, 741 (2011).
16. N. Denti, N. Ford, The courage to change the rules: A proposal for an essential health R&D treaty. PLoS Med. 2, e145 (2005).
17. S. J. Hoffman, J.-A. Røttingen, Assessing implementation mechanisms for an international agreement on research and development for health products. Bull. World Health Organ. 90, 854–863 (2012).
18. L. O. Gostin, Meeting basic survival needs of the world’s least healthy people toward a framework convention on global health (SSRN Scholarly Paper ID 1014082, Social Science Research Network, Rochester, NY, 2007) https://papers.ssrn.com/abstract=1014082 (accessed 18 July 2014).
19. United Nations, United Nations treaty collection (2022). https://treaties.un.org/ (accessed 19 October 2016).
20. E. M. Hafer-Burton, D. G. Victor, Y. Lupu, Political science research on international law: The state of the field. Am. J. Int. Law 106, 67–97 (2012) (accessed 19 October 2016).
21. G. Shaffer, J. Ginsburg, The empirical turn in international legal scholarship. Am. J. Int. Law 106, 1–55 (2012).
22. B. Koremenos, The Continent of International Law: Explaining Agreement Design (Cambridge University Press, Cambridge, UK, illustrated ed., 2016).
