European Commission

Demographic and health outcomes by Degree of Urbanisation

Perspectives from a new classification of urban areas

Written by
J. Vernon Henderson
Vivian Liu
Cong Peng
Adam Storeygard
November – 2019

The London School of Economics and Political Science

EUROPEAN COMMISSION

Directorate-General for Regional and Urban Policy
Directorate B — Policy
Unit B1 — Policy Development and Economic Analysis

Contact: Moray Gilland

E-mail: REGIO-B1-HEAD-OF-UNIT@ec.europa.eu

European Commission
B-1049 Brussels

Demographic and health outcomes by Degree of Urbanisation

Perspectives from a new classification of urban areas

Directorate-General for Regional and Urban Policy

2020 EN

Manuscript completed in November 2019

Luxembourg: Publications Office of the European Union, 2020

ISBN 978-92-76-15239-2
doi: 10.2776/37532

Demographic and health outcomes by Degree of Urbanisation: perspectives from a new classification of urban areas*

J. Vernon Henderson, London School of Economics†
Vivian Liu, London School of Economics‡
Cong Peng, University of California, San Diego§
Adam Storeygard, Tufts University¶

November 2019

Abstract

In practice, cities are typically defined using administrative and other criteria that are highly heterogeneous across countries. In this monograph, we leverage the Degree of Urbanisation concept from the Global Human Settlements Layer (GHS-SMOD) to yield a globally consistent definition of cities, towns and surburbs, and rural areas that is based on density and settlement size. Matching GHS-SMOD designations to geo-coded Demographic Health Survey (DHS) data, we explore differentials across space in a number of highly relevant policy indicators in utilities, health outcomes, educational attainment, fertility, and violence. Do cities offer a better environment for these outcomes? We report raw differences in outcomes between cities, towns, and rural areas circa 2015 for Sub-Saharan Africa, South Asia, Latin America, and Southeast Asia. With a few exceptions in health outcomes related to affluence and air pollution, outcomes improve going from rural areas to towns and suburbs, and then to cities. We restrict our sample to South Asia and Sub-Saharan Africa for in-depth analysis. We find that most outcomes improve over time, some dramatically. Even after accounting for sorting based on household and individual characteristics, or differences by political and administrative status, we find that most city-rural differentials persist, some key ones strongly.

*We thank Ying Chen for excellent research assistance, and Anna Aizer, Lewis Dijkstra, and Andrew Foster for helpful conversations. We gratefully acknowledge funding from the European Union, tender No 2018.CE.16.BAT.051 Demography and health by degree of urbanisation.

†J.V.Henderson@lse.ac.uk
‡V.Liu@lse.ac.uk
§C.Peng7@lse.ac.uk
¶adam.storeygard@tufts.edu

Executive Summary

In the developing world, evidence has long shown that city residents have higher incomes than their rural counterparts, as well as higher quality of life in many dimensions, including more access to public utilities and schools, better health, and lower fertility. This monograph explores such urban-rural differentials in a very wide set of measures in a large set of countries in Sub-Saharan Africa and South Asia, and a more limited sample of mostly poor countries in Latin America and Southeast Asia, to offer a fresh and more nuanced perspective on this general finding. To offer a better assessment, it also harnesses a new dataset delineating urban and rural areas consistently across countries based on the size and density of settlements rather than administrative definitions. Country coverage of the latter two regions is limited in the main data source, the Demographic and Health Surveys (DHS), so the focus is on Sub-Saharan Africa and South Asia.

We look at five sets of outcomes: utilities (electricity, water, and sanitation); health outcomes including infant mortality, vaccinations, and various indicators of morbidity; educational attainment of children; fertility including use of contraception and views on ideal numbers of children; and finally actual incidence of and attitudes towards violence and crime.

We use the Degree of Urbanisation concept, as implemented in the Global Human Settlements Layer (GHS-SMOD), to consistently define cities, towns and suburbs, and rural areas, based on density and built cover of 1-km grid cells as well as overall settlement size. We compare outcomes across these three types of places. Having consistent definitions across countries ensures that we are comparing the same settlement typologies, as opposed to administrative definitions, which differ quite substantially across countries in practice.

The monograph primarily reports raw differences in all outcomes between cities, towns and rural areas for the four regions circa 2015. For most outcomes, in most regions, outcomes are better in cities than in towns/suburbs which in turn are better than those in rural areas. One important exception is that rural areas in South Asia generally have better outcomes than towns/suburbs. Another is that health outcomes often associated with air pollution (cough, asthma) and affluence (obesity, diabetes, high blood pressure) are typically worse in cities than in rural areas. A final exception is crime.

We also look at differences over time between circa 2000 and circa 2015 for a subset of outcomes recorded consistently in both time periods. This part of the analysis is limited to a subset of Sub-Saharan African and South Asian countries, due to a lack of spatially explicit survey data circa 2000 for the other regions. We find improvements over time in nearly all categories, some quite dramatic, in both urban and rural areas, with substantial rural catchup in South Asia. The main exception is piped water, the provision of which has not kept up with population growth in urban Africa.

Differences in outcomes between cities and the countryside could exist for a number of reasons. First, differences could exist because of sorting of different types of individuals and households across zones. For example, cities have more educated populations, who may demand better school attendance and achievement by their children, who seek more and better health care, and who may have different attitudes towards, for example, domestic violence, family planning and the use of health care facilities. By controlling at the individual and household level for education, gender, and age, we can capture a considerable degree of the impact of sorting on city-rural differentials.

Second, differences could exist because of pure supply considerations such as spatial differences in public and private provision and quality of utilities, health facilities, and schools, with a bias towards cities. Third, differences could reflect broader features of the urban environment: air quality for health, the opportunity cost of raising children, stress, and local culture affecting attitudes and social norms concerning household violence, child rearing, family planning, and diet. Finally differences could exist because of political biases that influence the supply side, but also reflect more effective political influence of richer and more educated urban residents. These influences cannot be distinguished easily by the data at our disposal in this context, though we do provide limited evidence on the differentials between political/administrative centres and other cities.

In the last two sections, we look at the impact of sorting based on household and individual characteristics (Sub-Saharan Africa and South Asia), and differences by political and administrative status (Africa only) in the 2015 cross section.

Controlling for demographic characteristics that might affect family demand generally decreases city-rural differentials, by anywhere from trivial amounts to nearly half, with the largest impacts on health and fertility. Cities are still better places even for families with, for example, similar education levels.

Some city-rural differentials also persist after controlling for political and administrative regional status. However town effects generally disappear. National capitals have especially large differentials in several indicators. But being in a city seems to offer something beyond the political or administrative status of the area. This highlights some potential benefits of urbanisation in these countries, and suggests that better amenities draw people into cities. However it also suggests that if the supply of infrastructure and basic services in rural areas improved, this might slow migration, especially where city governments are overwhelmed with rapid population growth.

Detailed cross-sectional findings

What do we find in the 2015 cross section for our four regions in the raw data in terms of differences in outcomes among cities, towns and rural areas? For Sub-Saharan Africa and South Asia, what is the role of sorting in explaining these differentials?

In all regions, households are much better served by utilities in cities than in the countryside, and differences are often huge. For example, in Sub-Saharan Africa urban households are three times more likely to have access to safe water and electricity, and twice as likely to have improved sanitation, as their rural counterparts. Sorting plays a minimal role in these differences in both Sub-Saharan Africa and South Asia, suggesting that they are public supply driven.

Health outcomes provides a sharp contrast to utilities. While infant mortality, diarrhoea among children and vaccinations follow the expected pattern of being better in cities than the countryside, the differences in all four regions are modest, and for Sub-Saharan Africa and South Asia sometimes insignificant. Moreover sorting has a strong effect, significantly reducing or eliminating rural-city differences in many cases. It appears that efforts in many countries to provide basic health care everywhere have been more successful and less costly than efforts to equalize quality of utilities.

A second set of health outcomes shows a consistent distinct penalty to living in cities instead of the countryside. These include diseases often associated with affluence (adult obesity and high blood pressure, adult female diabetes) and with air pollution (child cough and adult female asthma), where only the first in each category are available worldwide, with the remainder limited to India. Sorting has little impact on these morbidity differences, but does indicate that to some degree these are diseases of the middle and upper classes, consistent with diets higher in sugar, fat and unprocessed grains, and lack of exercise.

Schooling attainment (proportion of 16-year-olds who have completed at least 8 grades) is much higher in cities in all regions, except South Asia. Sorting matters but not enormously, suggesting that public policies affecting supply are likely important, as for utilities. Interestingly in Sub-Saharan Africa, part of the reason children do not complete grade 8 by age 16 is that many 12- and 13-year-old children have still not completed even grade 4. They attend but do not finish grades. However a large portion of these older children persist in attendance, so that they have completed grade 4 by age 16.

Actual and desired fertility rates are lower in cities than the countryside in all regions, although modestly so in South Asia. Correspondingly, use of modern contraception is higher in cities, except for South Asia where it is surprisingly lower. Sorting explains some but not all of these differences.

Finally, we explore two dimensions of violence and crime. First on domestic violence, fewer urban residents find wife beating justifiable, and correspondingly fewer wives have experienced violence. Sorting plays a very small role in explaining this in Sub-Saharan Africa, but a larger one in South Asia. Second for crime more generally, we find that in contrast to domestic violence, households in cities are more likely to experience physical attacks and home robberies, and slightly more likely to have feared crime in their own home or felt unsafe walking in their neighbourhood. As with domestic violence, sorting does not explain these city-rural differentials.

Focusing on Sub-Saharan Africa, we further take political and administrative regional status into account, looking at the role played by being in a national or regional capital, administrative designation as urban by the national census, and local population density. After including these controls, though town-rural differentials generally disappear, city-rural differentials generally persist and spatial patterns do not change. We find that national capitals and administrative urban status are strongly related to our outcomes in utilities, health, schooling, fertility, and contraception. These results are consistent with the possibility of a national capital (and more general urban) bias, especially in the provision of utilities and schooling.

Abstract

Les villes sont généralement définies à partir de critères, notamment administratifs, qui varient grandement d'un pays à un autre. Dans cette étude, nous utilisons le concept de taux d'urbanisation, tel que défini par le « Global Human Settlements Layer » (GHS-SMOD), afin de proposer une définition des villes, villages, banlieues et zones rurales centrée sur leur étendue et densité de population. En recoupant les outils conceptuels proposés par le GHS-SMOD et les données géocodées du « Demographic Health Survey » (DHS), nous explorons les différences entre ces espaces en termes de services publics, conditions de santé, niveau d'instruction, de fécondité et de violence. Les villes offrent-elles un environnement comparativement plus favorable au regard de ces indicateurs ? Dans cette étude, nous constatons des différences importantes entre les villes, les villages et les zones rurales d'Afrique subsaharienne, d'Asie du Sud, d'Amérique latine et d'Asie du Sud-Est aux alentours de 2015. A l'exception de quelques effets négatifs liés à la pollution atmosphérique et à l'abondance de ressources sur les conditions de santé, les indicateurs considérés s'améliorent à mesure que l'on passe des zones rurales aux villages, puis aux banlieues et enfin aux villes. En restreignant notre échantillon de recherche à l'Asie du Sud et à l'Afrique subsaharienne, que nous analysons en profondeur, nous constatons que la plupart des indicateurs considérés s'améliorent avec le temps, certains de manière significative. Mais même après avoir pris en compte les différences liées à l'auto-sélection des ménages, dues à des caractéristiques individuelles ou à des différences régionales de statut politique et administratif, nous constatons que la plupart des différences entre villes et zones rurales persistent.

Résumé

Dans les pays en développement, des études ont depuis longtemps signalé que les résidents des villes ont de plus hauts revenus que ceux des zones rurales, une meilleure qualité de vie, notamment en termes de services publics, d'accès à l'éducation, de conditions de santé, et un taux de fécondité plus faible. Cette étude se penche sur ces différences entre zones rurales et urbaines telles que rapportées par un grand nombre de mesures dans un large panel de pays en Afrique subsaharienne et Asie du Sud, et, dans une moindre mesure, dans des pays pauvres d'Amérique latine et Asie du Sud-Est, afin d'apporter une nouvelle perspective, plus nuancée. Afin de proposer une meilleure évaluation de ces différences, notre étude exploite de nouvelles données délimitant les zones urbaines et rurales de la même manière d'un pays à un autre, en fonction de l'étendue et densité des établissements humains dans la région considérée, plutôt que les définitions administratives habituelles. En Amérique latine et Asie du Sud-Est, les données dont nous disposons – grâce au « Demographic Health Survey » (DHS) – sont limitées. Notre attention se porte donc principalement sur l'Afrique subsaharienne et l'Asie du Sud.

Nous nous intéressons à cinq indicateurs : les services publics (distribution d'électricité, d'eau et assainissement), les conditions de santé (y compris la mortalité infantile, le taux

de vaccination et différents indicateurs de morbidité), le niveau d'instruction des enfants, le taux de fécondité (y compris l'utilisation de contraceptifs et le nombre idéal d'enfants selon la population) et, enfin, la fréquence de la violence et le taux de criminalité ainsi que l'attitude populaire envers ces phénomènes.

Nous utilisons le concept de taux d'urbanisation, tel que proposé dans le cadre du « Global Human Settlements Layer » (GHS-SMOD), afin de définir de manière systématique les villes, villages et banlieues, ainsi que les zones rurales, en fonction de leur étendue, densité et de la surface construite par maille de 1km2. Nous comparons les résultats obtenus dans chacun de ces trois types d'espaces. Utiliser les mêmes définitions appliquées à différents pays nous permet de comparer la même typologie d'établissement humain, à la différence des définitions administratives, qui, en pratique, varient fortement entre les pays.

Cette étude présente principalement les différences brutes entre villes, villages et zones rurales pour les quatre régions considérées aux alentours de 2015, quel que soit l'indicateur considéré. Pour la plupart de ces indicateurs, et dans la plupart des régions, les résultats sont meilleurs et l'environnement semble donc plus favorable dans les villes que dans les villages et banlieues, dont l'environnement reste cependant plus favorable que celui des zones rurales. Les zones rurales d'Asie du Sud constituent une exception, l'environnement y étant généralement plus favorable que dans les villages et banlieues. Une autre exception se situe au niveau des effets négatifs de la pollution atmosphérique (toux, asthme) et de l'abondance (obésité, diabète, hypertension) sur la santé, qui sont généralement plus importants en ville que dans les zones rurales. Enfin, le taux de criminalité est aussi une exception importante.

Nous examinons également les différences dans le temps, entres les années 2000 et 2015, pour un sous-ensemble de résultats enregistrés de la même façon les deux années. Cette partie de l'analyse se limite à un sous-ensemble de pays d'Afrique subsaharienne et d'Asie du Sud, en raison du manque de données spatialement explicites pour les autres régions dans les années 2000. Nous constatons des améliorations au fil du temps de presque tous les indicateurs considérés, certaine fois de manière significative, tant dans les zones urbaines que rurales, avec un développement important des zones rurales d'Asie du Sud notamment. La principale exception est l'eau courante, dont l'approvisionnement n'a pas suivi la croissance démographique en Afrique urbaine.

Les différences entre ville et campagne peuvent exister pour de nombreuses raisons. D'abord, des différences peuvent apparaître du fait de l'auto-sélection des individus et ménages dans les deux zones considérées. Par exemple, les villes ont une population en moyenne plus éduquées, qui exige de ses enfants une assiduité et des résultats plus élevés, qui demande un meilleur accès aux soins, et qui tend à avoir une attitude différente envers, par exemple, la violence domestique, la contraception et le recours aux services de santé. En contrôlant le niveau d'éducation, le genre et l'âge au niveau des individus et des ménages, nous sommes en mesure de saisir l'impact de l'auto-sélection sur les différences entre villes et zones rurales.

Deuxièmement, des différences peuvent s'établir du fait de variation dans l'offre, telles que les différences spatiales entre prestations publiques et privées, quant à l'accès aux soins, à l'éducation, avec un biais en faveur des villes. Troisièmement, des différences peuvent refléter des caractéristiques plus générales de l'environnement urbain : la qualité de l'air, le coût d'opportunité de l'éducation des enfants, le stress et la culture locale affectant les attitudes et les normes sociales concernant la violence domestique, l'éducation des enfants, la planification familiale et l'alimentation. Enfin, des différences peuvent exister en raison de biais politiques qui non seulement influencent l'offre mais reflètent aussi la grande influence politique des plus riches et des plus éduqués des résidents des villes. Ces influences peuvent difficilement être distinguées dans les données dont nous disposons, bien que l'on apporte des preuves (certes limitées) des différences entre les grands centres politiques et administratifs urbains et les autres villes.

Dans les deux dernières sections de cette étude, nous analysons l'impact de l'auto-sélection liée à des caractéristiques individuelles et à certains types de ménages (en Afrique subsaharienne et en Asie du Sud), ainsi que celle liée à des différences de statut politique et administratif (en Afrique seulement), dans l'échantillon de 2015. La prise en compte des caractéristiques démographiques susceptibles d'affecter les demandes des familles réduit généralement l'écart entre zones urbaines et rurales, certain fois marginalement mais dans d'autre cas de presque la moitié, avec les effets les plus importants dans le cas de la santé et de la fécondité. Les villes restent de meilleurs environnements pour les familles ayant, par exemple, un niveau d'éducation similaire.

Certaines différences entres zones urbaines et rurales persistent même après avoir pris en compte les différences liées au statut politique et administratif des régions. Cependant, ces effets disparaissent généralement au niveau des villages. Les capitales nationales ont un avantage parfois considérable au regard de plusieurs indicateurs. Mais vivre en ville semble offrir quelque chose qui va au-delà du simple statut politique ou administratif de la zone. Cela met en évidence certains bénéfices potentiels de l'urbanisation dans ces pays, et suggère que de meilleures infrastructures attirent la population vers les villes. Cependant, cela suggère aussi que si l'offre d'infrastructure et de services de base dans les zones rurales était améliorée, cela pourrait ralentir la migration, en particulier dans les villes où les autorités sont dépassées par une croissance rapide de la population.

Résultats détaillés de notre étude transversale

Que nous révèlent les données brutes de l'échantillon de 2015 de nos quatre régions au sujet des différences entre villes, villages et zones rurales, au vu des cinq indicateurs considérés ? Dans le cas de l'Afrique subsaharienne et de l'Asie du Sud, quel rôle joue l'auto-sélection individuelle dans l'explication de ces différences ?

Dans toutes les régions considérées, les ménages ont un meilleur accès aux services publics en ville qu'à la campagne, et les différences sont souvent énormes. Par exemple, dans les ménages citadins d'Afrique subsaharienne ont trois fois plus de chance d'avoir accès à de l'eau potable et à l'électricité, et deux fois plus de chance de bénéficier d'un accès à d'installations sanitaires de qualité, que les ménages ruraux. L'auto-sélection joue un rôle minime dans ces différences, que cela soit en Afrique subsaharienne ou en Asie du Sud, ce qui suggère que celles-ci sont, avant tout, liées à l'offre publique.

Les résultats obtenus sur le plan de la santé contrastent nettement avec ceux obtenus au sujet des services publics. Alors que les taux de mortalité infantile, de diarrhées chez les enfants et de vaccination sont, comme on pouvait s'y attendre, meilleurs dans les villes que dans les zones rurales, ces différences sont en réalité modestes dans les quatre régions et parfois insignifiantes en Afrique subsaharienne et en Asie du Sud. De plus, il apparaît que l'auto-sélection a un effet important, réduisant significativement voire éliminant les différences entre villes et zones rurales dans bien des cas. Les efforts faits dans plusieurs pays pour offrir une couverture santé et un accès aux soins de base partout semblent avoir été plus efficaces et moins coûteux que les efforts visant à égaliser la qualité des services publics.

Une deuxième série de résultats en matière de santé montre qu'il existe un désavantage à vivre en ville plutôt qu'à la campagne. Il s'agit notamment de maladies liées à l'abondance de ressources (obésité chez l'adulte et hypertension, diabète chez la femme adulte) et à la pollution atmosphérique (toux chez l'enfant et asthme chez la femme adulte). Le premier de chaque catégorie est visible dans le monde entier, le second étant limité à l'Inde. L'auto-sélection des individus n'a qu'un faible impact sur ces différences de morbidité, mais indique que ces maladies affectent plutôt les classes moyennes et les plus riches, ce qui est logique étant donné leur régime riche en sucre, gras et céréales non-transformées, et leur manque d'exercice.

Le niveau d'instruction des enfants (mesuré à partir de la proportion de jeunes de 16 ans ayant atteint la troisième – « 8th grade » dans le système anglo-saxon) est bien plus élevé dans les villes de toutes les régions considérées, hormis en Asie du Sud. L'auto-sélection joue un rôle ici aussi mais seulement mineur, ce qui suggère que les politiques publiques affectant l'offre en termes d'éducation sont importantes, de la même manière que pour les services publics. Il est intéressant de remarquer qu'en Afrique subsaharienne, une des raisons pour lesquelles les enfants n'ont généralement pas atteint la troisième à l'âge de 16 ans est que de nombreux jeunes de 12 et 13 ans n'ont toujours pas terminé leur classe de CM1 (« 4th grade »). Ils suivent les cours mais ne terminent pas le programme. Cependant, une large proportion de ces jeunes continue d'aller en classe, et finit par terminer le programme de CM1 avant l'âge de 16 ans.

Les taux de fécondité réels et souhaités sont plus faibles dans les villes que dans les zones rurales de toutes les régions considérées, bien que de manière modeste en Asie du Sud. L'utilisation de contraceptifs modernes est plus élevée en ville, à l'exception de l'Asie du Sud, où celle-ci est étonnamment plus faible. L'auto-sélection des individus explique certaines de ces différences mais pas toutes.

Nous explorons, en dernier lieu, deux facettes de la violence et de la criminalité. Tout d'abord, en ce qui concerne la violence domestique, il apparaît qu'une plus faible proportion de citadins trouve que battre sa femme est justifiable, ce qui se traduit par un plus faible nombre de victimes de violence domestiques qu'en zones rurales. L'auto-sélection ne joue qu'un rôle mineur dans l'explication de cette différence en Afrique subsaharienne, alors qu'elle a un effet plus important en Asie du Sud. Au sujet de la criminalité de manière plus générale, nous constatons que, contrairement aux violences domestiques, les ménages des villes sont plus susceptibles d'être victimes d'agressions physiques et de cambriolages. Ils sont aussi légèrement plus susceptibles d'avoir craint qu'un crime soit commis chez eux ou d'avoir eu un sentiment d'insécurité en se promenant dans leur quartier. Comme dans le cas des violences domestiques, l'auto-sélection n'explique pas cette différence entre villes et zones rurales.

En ce qui concerne l'Afrique subsaharienne en particulier, nous prenons également en compte le statut politique et administratif régional en cherchant à évaluer le rôle joué par le statut de capitale nationale ou régionale, la désignation administrative des zones urbaines par le recensement national et la densité de population locale. Après prise en compte de ces éléments, bien que les différences entre villages et zones rurales disparaissent généralement, les différences entre villes et zone rurales persistent et les caractéristiques spatiales ne varient pas. Il apparaît que les capitales nationales et les centres administratifs urbains bénéficient d'un meilleur accès aux services publics, à l'éducation et aux soins tout en ayant un taux plus faible de fécondité et une utilisation plus importante de moyens de contraception. Ces résultats sont en adéquation avec un possible biais en faveur des capitales nationales (et des villes plus généralement), notamment en termes de prestation des services publics et d'accès à l'éducation.

Zusammenfassung

In der Praxis werden Städte typischerweise anhand administrativer oder ähnlicher Kriterien abgegrenzt, die international nur eingeschränkt vergleichbar sind. In diesem Beitrag nutzen wir das Konzept des Urbanisierungsgrads aus dem Global Human Settlements Layer (GHS-SMOD), um eine weltweit einheitliche, auf Bevölkerungsdichte und Siedlungsgröße basierende Definition für Großstädte, Städte und Vororte, sowie ländliche Gegenden zu generieren. Durch die Verknüpfung der auf dem GHS-SMOD basierenden Abgrenzungen mit geo-kodierten Daten des Demographic and Health Surveys (DHS) können wir räumliche Unterschiede hinsichtlich politisch hochrelevanter Indikatoren im Bereich der öffentlichen Infrastruktur, dem Gesundheitswesen, dem Bildungsniveau, dem Geburtenverhalten und der Verbreitung von Gewalt darstellen. Fallen diese Indikatoren in Großstädten besser aus? Für ca. 2015 stellen wir zunächst unbereinigte Unterschiede zwischen Großstädten, Städten und ländlichen Gegenden in Subsahara-Afrika, Südasien, Lateinamerika und Südostasien dar. Abgesehen von einigen mit Wohlstand und Luftverschmutzung assoziierten Krankheiten fallen unsere Indikatoren in Großstädten besser aus als in Städten und Vororten, und in letzteren besser als in ländlichen Gegenden. Für eine Detailanalyse begrenzen wir unsere Stichprobe auf Südasien und Subsahara-Afrika. Wir zeigen, dass sich die meisten Indikatoren über die Zeit verbessern, in einigen Fällen sogar dramatisch. Selbst bei expliziter Berücksichtigung der räumlichen Selektion von Haushalten und Individuen, und nach Berücksichtigung des jeweiligen politisch-administrativer Status großer Städte finden wir, dass die meisten Stadt-Land-Unterschiede weiterhin Bestand haben – in einigen wichtigen Fällen in bedeutendem Ausmaß.

Kurzfassung

Für Entwicklungsländer ist schon länger belegt, dass Stadtbewohner im Vergleich zu Landbewohnern ein höheres Einkommen erzielen und eine höhere Lebensqualität genießen, was sich beispielsweise durch einfacheren Zugang zu öffentlicher Infrastruktur und Schulen, bessere Gesundheit und niedrigere Geburtenraten ausdrückt. Dieser Beitrag entwirft anhand einer breiten Palette an Indikatoren eine aktuelle, nuancierte Perspektive auf derartige Stadt-Land-Unterschiede, die für eine große Gruppe von Ländern in Subsahara-Afrika und Südasien, sowie einige vornehmlich arme Länder in Lateinamerika und Südostasien analysiert werden. Um die Analysequalität zu erhöhen, nutzen wir einen neuen Datensatz, mittels dessen die Abgrenzung von städtischen und ländlichen Regionen über alle Untersuchungsländer hinweg konsistent erfolgt – basierend auf der Siedlungsgröße und der Bevölkerungsdichte statt auf administrativen Definitionen. Die Menge der Untersuchungsländer in den beiden letztgenannten Regionen ist begrenzt durch unsere Hauptdatenquelle, die Demographic and Health Surveys (DHS); unser Fokus liegt daher auf Subsahara-Afrika und Südostasien.

Wir betrachten fünf Indikatorengruppen. Diese messen den Zugang zu Versorgungsinfrastruktur (Strom-, Wasser- und Abwassersysteme), Gesundheitsindikatoren (darunter

Kindersterblichkeit, Impfungshäufigkeit und diverse Morbiditätsindikatoren), das Bildungsniveau von Kindern, das Fertilitätsverhalten (auch hinsichtlich der Verwendung von Verhütungsmitteln und der erwünschten Kinderzahl) und schließlich die Häufigkeit und Akzeptanz von Gewalt und Kriminalität.

Wir verwenden das im Global Human Settlements Layer (GHS-SMOD) definierte Konzept des Urbanisierungsgrads um Großstädte, Städte und Vororte, und ländliche Gegenden konsistent voneinander abzugrenzen, wobei Daten zur Gesamtsiedlungsgröße, sowie zur Bevölkerungsdichte und bebauten Fläche in 1-km-Gitternetzeinheiten zur Anwendung kommen. All unsere Indikatoren werden zwischen diesen drei Ortstypen verglichen. Durch Nutzung länderübergreifend konsistenter Definitionen stellen wir sicher, dass unabhängig vom Untersuchungsland stets die gleichen Siedlungstypologien miteinander vergleichen werden. Verwaltungsrechtliche, in der Praxis je nach Land variierende Abgrenzungen beeinflussen unsere Ergebnisse daher nicht.

Dieser Beitrag stellt zunächst grundsätzliche Unterschiede für alle Indikatoren zwischen Großstädten, Städten und ländlichen Gegenden in den vier untersuchten Regionen für ca. 2015 dar. Für die meisten Indikatoren und die meisten Regionen gilt, dass die Indikatoren in Großstädten günstiger als in Städten und Vororten, und dort wiederum günstiger als in ländlichen Gegenden ausfallen. Eine wichtige Ausnahme bilden ländliche Gebiete in Südasien, die allgemein besser abschneiden als Städte und Vororte derselben Region. Eine weitere Ausnahme bilden einige mit Luftverschmutzung (Husten, Asthma) und Wohlstand (Fettleibigkeit, Diabetes, Bluthochdruck) assoziierte Gesundheitsindikatoren, die in Großstädten typischerweise ungünstiger ausfallen als in ländlichen Gegenden. Außerdem weichen unsere Kriminalitätsindikatoren vom grundsätzlichen Muster ab.

Für einige zwischen ca. 2000 und ca. 2015 konsistent erhobene Indikatoren untersuchen wir deren zeitliche Veränderung. Dieser Teil der Analyse ist auf eine Teilgruppe von Ländern in Subsahara-Afrika und Südasien begrenzt, mangels explizit geo-kodierter Studiendaten für andere Regionen in den 2000ern. Über die Zeit hinweg verbessern sich nahezu alle Indikatoren – einige davon durchaus dramatisch – sowohl in städtischen als auch in ländlichen Gegenden, wobei insbesondere die ländlichen Regionen Südasiens merklich aufholen. Die wichtigste Ausnahme stellt die Leitungswasserversorgung dar, die in den städtischen Regionen Afrikas nicht mit dem Bevölkerungswachstum Schritt gehalten hat.

Die unterschiedliche Ausprägung unserer Indikatoren in Städten und ländlichen Gegenden könnte verschiedene Ursachen haben. Zunächst könnte es sein, dass sich Einzelpersonen und Haushalte mit unterschiedlichen Merkmalen systematisch auf städtische und ländliche Gegenden verteilen (räumliche Selektion). Beispielsweise weisen Großstädte eine gebildetere Bevölkerung auf, die im Vergleich zu Landbewohnern wohlmöglich höheren Wert auf Schulanwesenheit und -leistung ihrer Kinder legt, eine bessere Gesundheitsversorgung fordert, und andere Einstellungen zu häuslicher Gewalt, Familienplanung und der Nutzung von Gesundheitseinrichtungen aufweist. Indem wir auf Individual- und Haushaltsebene für das Bildungsniveau, Geschlecht und Alter kontrollieren, können wir den Einfluss solcher Selektionseffekte größtenteils einfangen.

Zweitens könnten die Stadt-Land-Differenzen schlicht eine räumliche Ungleichverteilung des lokalen Angebots widerspiegeln, etwa hinsichtlich der öffentlichen und privaten Bereitstellung und Qualität von Versorgungsinfrastruktur, Gesundheitseinrichtungen und Schulen, die in Großstädten wahrscheinlich dichter ist. Drittens könnten die gemessenen Unterschiede auf allgemeinere Besonderheiten eines städtischen Umfelds zurückzuführen sein: Die städtische Luftqualität und deren Einfluss auf die Gesundheit; die Opportunitätskosten des Kinderkriegens und -erziehens; Stress; die lokale Kultur und ihr Einfluss auf Werte und soziale Normen, etwa hinsichtlich häuslicher Gewalt, Kindererziehung, Familienplanung und Ernährungsgewohnheiten. Schließlich könnten Stadt-Land-Unterschiede auch politische Entscheidungen über die Verteilung des Versorgungsangebots widerspiegeln, das aufgrund des stärkeren politischen Einflusses wohlhabenderer und höher gebildeter urbaner Bevölkerungsschichten systematisch zugunsten von Städten verzerrt sein könnte. Diese Effekte können anhand der für unsere Untersuchung relevanten Datensätze nicht ohne Weiteres isoliert werden. Wir zeigen jedoch, dass sich politisch-administrative Zentren hinsichtlich unserer Indikatoren von anderen Großstädten unterscheiden.

In den letzten beiden Abschnitten dieses Beitrags nutzen wir Querschnittsdaten für 2015 um den Einfluss räumlicher Selektion nach Haushalts- und Individualmerkmalen (Subsahara-Afrika und Südasien), sowie die möglicherweise aus der politisch-administrativen Sonderstellung einzelner Städte resultierende Vorteile zu untersuchen (nur Afrika).

Die explizite Berücksichtigung demographischer, die Familiennachfrage beeinflussender Charakteristika reduziert Stadt-Land-Unterschiede durchgängig – teils nur geringfügig, in anderen Fällen fast um die Hälfte. Insbesondere in Hinsicht auf Gesundheitsindikatoren und das Geburtenverhalten nehmen die Unterschiede deutlich ab. Dennoch geht es großstädtischen Familien im Vergleich zu ländlichen Familien mit ähnlichen Charakteristika (zum Beispiel hinsichtlich des Bildungsniveaus) im Schnitt besser.

Einige Stadt-Land-Unterschiede bleiben selbst nach statistischer Kontrolle für den besonderen Status regionaler Politik- und Verwaltungszentren bestehen. Indes lösen sich die Unterschiede zwischen kleineren Städten und ländlichen Gegenden auf. Nationale Hauptstädte weisen besonders große Unterschiede zu anderen Großstädten, Städten und ländlichen Gegenden auf. Dennoch scheint sich das Leben in einer Großstadt über die aus deren politischer oder administrativer Bedeutung resultierenden Vorteile hinaus auszuzahlen. Das Bestehen einer Restdifferenz zwischen Stadt und Land deutet einerseits darauf hin, dass die untersuchten Länder von der Urbanisierung profitieren und dass die Annehmlichkeiten urbanen Lebens Menschen in die Großstädte ziehen. Andererseits bestehen klare Hinweise darauf, dass eine Verbesserung der Infrastruktur und Versorgung mit grundlegenden Dienstleitungen in ländlichen Gegenden die Land-Stadt-Migration abschwächen würde, vor allem dort, wo städtische Behörden durch rasantes Bevölkerungswachstum überfordert sind.

Detaillierte Querschnittsergebnisse

Welche Unterschiede zwischen Großstädten, Städten und ländlichen Gegenden werden sichtbar, wenn wir unsere Indikatoren simultan für alle vier Regionen in einer Querschnittsanalyse für das Jahr 2015 untersuchen? Und in welchem Maße kann räumliche Selektion diese Unterschiede in Subsahara-Afrika und Südasien erklären?

In allen vier Regionen genießen in Städten lebende Haushalte eine bessere Versorgung mit öffentlicher Infrastruktur als auf dem Land, oft mit beträchtlichen Unterschieden. Beispielsweise haben städtische Haushalte in Subsahara-Afrika im Vergleich zu ähnlich konstituierten ländlichen Haushalten mit dreimal höherer Wahrscheinlichkeit Zugang zu sauberem Wasser und Elektrizität, sowie mit doppelter Wahrscheinlichkeit Zugang zu sanitären Einrichtungen. Räumliche Selektion spielt für die Erklärung dieser Unterschiede in Subsahara-Afrika und Südasien nur eine geringfügige Rolle. Stattdessen scheint die ungleiche lokale Verfügbarkeit entsprechender Angeboten die Unterschiede zu erklären.

Ein stark davon abweichendes Bild liefert die Analyse von Gesundheitsindikatoren. Zwar entsprechen die Differenzen bezüglich der Kindersterblichkeit, der Häufigkeit von Durchfallerkrankungen bei Kindern und des Impfverhaltens den erwarteten Mustern, fallen also in Städten günstiger aus als auf dem Land. In allen vier Regionen sind die Unterschiede jedoch gering, in Subsahara-Afrika und Südasien teilweise sogar insignifikant. Darüber hinaus hat räumliche Selektion einen starken Einfluss und reduziert oder beseitigt die Stadt-Land-Unterschiede in vielen Fällen gänzlich. Dies deutet darauf hin, dass bisherige Bemühungen um eine flächendeckende Gesundheitsversorgung in vielen Ländern erfolgreicher und weniger kostspielig waren als Versuche, die Qualität der Infrastruktur flächendeckend anzugleichen.

Andere Gesundheitsindikatoren fallen in Städten im Vergleich zu ländlichen Gegenden dagegen durchgehend nachteiliger aus. Dazu gehören Wohlstandskrankheiten (Fettleibigkeit bei Erwachsenen, Bluthochdruck, Diabetes bei erwachsenen Frauen) und mit Luftverschmutzung assoziierte Erkrankungen (Husten bei Kindern, Asthma bei erwachsenen Frauen), wobei nur der erstgenannte Indikator jeweils beider Kategorien weltweit untersucht werden kann, während sich eine Analyse weiterer Indikatoren auf Indien beschränken muss. Räumliche Selektion hat kaum Einfluss auf die Morbiditätsdifferenzen. Es wird jedoch deutlich, dass sich derartige Krankheiten vornehmlich auf die mittleren und oberen Gesellschaftsschichten konzentrieren – ein Befund, der auf eine zucker-, fett-, und an unverarbeitetem Getreide reichere Ernährung, sowie mangelnde körperliche Bewegung in diesen Bevölkerungsgruppen hinweist.

Das Schulbildungsniveau (gemessen als Prozentsatz aller 16jährigen mit mindestens acht Jahren Schulbesuch) fällt in Großstädten über alle Regionen hinweg viel höher aus, mit Ausnahme von Südasien. Räumliche Selektion spielt keine maßgebliche Rolle, was darauf hindeutet, dass bisherige politische Maßnahmen zur Ausweitung des Bildungsangebots wie im Fall der öffentlichen Infrastruktur wirkungsvoll sind. In Subsahara-Afrika erklärt sich das niedrige Schulbildungsniveau, beziehungsweise der geringe Anteil von 16jährigen mit mindestens acht Jahren Schulbildung interessanterweise durch die Tatsache, dass dort viele 12 oder 13 Jahre alte Kinder noch nicht einmal das vierte Schuljahr abgeschlossen haben. Diese Kinder gehen zwar zur Schule, schließen aber das Schuljahr nicht ab. Ein großer Anteil dieser Kinder geht dennoch durchgehend zur Schule, sodass mit 16 immerhin das vierte Schuljahr abgeschlossen wird.

Sowohl die tatsächlichen Geburtenraten als auch die erwünschte Familiengröße sind über alle Regionen hinweg in Großstädten niedriger als auf dem Land, in Südasien allerdings nur geringfügig. Dementsprechend fällt die Nutzungsrate moderner Verhütungsmittel in Großstädten höher aus, außer in Südasien, wo sie überraschenderweise geringer ist. Räumliche Selektion erklärt diese Stadt-Land-Unterschiede teilweise, aber nicht gänzlich.

Abschließend analysieren wir zwei Dimensionen der Gewalt und der Kriminalität. Bezüglich der Verbreitung häuslicher Gewalt zeigen wir, dass Stadtbewohner Gewalt gegen Ehefrauen seltener legitim finden und in Städten lebende Frauen entsprechend weniger Gewalt erfahren. Räumliche Selektion spielt in der Erklärung dieser Muster in Subsahara-Afrika eine nur geringfügige Rolle, in Südasien jedoch eine wichtige. Im Gegensatz zu unseren Befunden zur häuslichen Gewalt stehen jene zu allgemeineren Kriminalitätsindikatoren: Haushalte in Großstädten fallen mit einer höheren Wahrscheinlichkeit physischen Angriffen und Wohnungseinbrüchen zum Opfer und geben mit einer höheren Wahrscheinlichkeit an, sich vor Kriminalität im eigenen Heim oder bei der Bewegung durch ihre Nachbarschaft zu fürchten. Wie im Fall der häuslichen Gewalt können diese Stadt-Land-Unterschiede nicht durch räumliche Selektion erklärt werden.

Für Subsahara-Afrika werfen wir außerdem einen genaueren Blick auf etwaige Besonderheiten in regionalen Politik- und Verwaltungszentren. Gibt es eine Verbindung zwischen dem Status nationaler oder regionaler Hauptstädte, der verwaltungsmäßigen Klassifizierung als „städtisch“, der lokalen Bevölkerungsdichte und unseren Indikatoren? Bei Berücksichtigung dieser Merkmale verschwinden zwar die Stadt-Land-Unterschiede grundsätzlich, aber Unterschiede zwischen Großstädten und ländlichen Regionen bleiben weiterhin mit ähnlichen räumlichen Mustern bestehen. Wir zeigen abschließend, dass nationale Hauptstädte und verwaltungsrechtlich als städtisch klassifizierte Regionen deutlich abweichende Indikatoren im Bereich der öffentlichen Infrastruktur, Gesundheit, Ausbildung, Geburtenrate und Verhütung aufweisen. Diese Befunde weisen darauf hin, dass nationale Hauptstädte (und Städte grundsätzlich) hinsichtlich der Bereitstellung von öffentlicher Infrastruktur und Schulen politisch bevorzugt werden.

1 Introduction

Urban residents are better-off on average than their counterparts in rural areas in most countries in the developing world, according to a wide variety of measures (Gollin et al., 2018, 2013; Chauvin et al., 2017). Economists have long argued that much higher incomes in cities reflect the production benefits from agglomeration in rich countries, and recent work has demonstrated this urban-rural differential in income (e.g. Henderson et al., 2019; Quintero and Roberts, 2018; Combes et al., 2019), and correspondingly, poverty rates (Ferré et al., 2012) in developing countries. Other work has shown this urban advantage in a much broader set of outcomes including health and mortality (Cleland et al., 1992), and pollution and the provision of utilities and other public goods (Gollin et al., 2018; Dorélien et al., 2013).

However, quantifying rural-urban differences is complicated by a number of factors. Foremost are the challenges inherent to defining and measuring urban residence (Dorélien et al. 2013; Gollin et al. 2018; Balk et al. 2006). This adds considerable uncertainty to the magnitudes of these differentials. Surveys containing rich socioeconomic outcomes typically report urban residence based on a binary administrative definition. In some cases, such definitions are based on objective criteria related to size of a settlement and sometimes its density, or more rarely the non-agricultural share of the working population. However, the relevant thresholds differ across countries (United Nations, 2018) as do criteria for defining the extent of the urban area. In other cases, the criteria are explicitly ad hoc or more subjective such as the nature of the settlement and its centricity and infrastructure. For any set of criteria, there is an enormous diversity among urban areas along several dimensions. (Ferré et al., 2012; Corker, 2017).

Having consistency within and across countries in defining urban areas and their sizes in the developing world is critical in a context of rapid urbanisation. While, according to the UN, just 40% of the population of non-OECD countries lived in urban areas in 2000, 61% is expected to be urban in 2040, implying a total of 2.75 billion new urban residents (United Nations, 2018). Understanding what is urban, where people are moving, and who lives in high density areas has implications for where more infrastructure and services will be needed.

Conversely, continuing vast differences in service levels across space may affect migration and urbanisation rates.

For example in Sub-Saharan Africa, Gollin et al. (2018) argue that many public-service-related outcomes improve with density. A potentially related fact is that in Africa, excluding primate cities, the main occupation of urban residents as defined by country censuses is agriculture (Henderson and Kriticos, 2018). While farmers may choose to live in cities to work secondary occupations in the off-season and to provide non-farm job access for family members (Henderson et al., 2019), they may also choose to live in cities to have access to services with poor availability in rural areas. That is, service level differentials may in part drive urbanisation. Finally provision of services may discriminate in favour of national and regional political capitals, inducing in-migration to these particular cities.

One factor driving these service level differentials is that the per capita cost of providing many services, such as utilities and medical facilities, is decreasing in population density due to scale economies (Ahlfeldt and Pietrostefani, 2019). However, in a number of countries, equality in provision of services across space is an explicit national objective. Provision of course is different than uptake, and uptake may be related to the socio-economic status of households. Due to sorting, people in cities could have better amenity outcomes even with the same level of provision, based upon their capacities, demands and needs.

In this monograph, we study spatial differentials in a wide variety of socioeconomic indicators in 41 countries throughout Sub-Saharan Africa, Latin America, and South and Southeast Asia, using a new spatially explicit and globally consistent tripartite classification of settlements. In particular, the new classification, the Degree of Urbanisation, as implemented in the Settlement Grid of the Global Human Settlements Layer (GHS-SMOD) produced by the Joint Research Centre of the European Commission, combines administrative data on population density and the population size of contiguous settlements with a classification of built-up areas based on satellite data to allocate people in an administrative unit to 1 km grid squares. It separately identifies cities, as contiguous grid squares with density over 1500 per km^2 and total size over 50,000, and towns/suburbs, with density over 300 per km^2 and total populations over 5000. The Global Human Settlements Layer itself builds on previous work consistently delineating urban areas. Several datasets based primarily on remote sensing data have been produced since 2001.[1] For example, the Global Rural-Urban Mapping Project (GRUMP; Balk et al. 2006; CIESIN, IFPRI, World Bank, and CIAT, 2011) combined a coarse remote-sensing indicator of urban status, night-time lights, with census-based measures of settlement size and density. GHS uses substantially higher resolution data on built-up area, at 30-meters or finer, applying globally consistent size and density thresholds, and reporting for 4 time periods over 40 years.

We study highly policy-relevant indicators related to utilities, health, education, violence, and fertility, each with data that are relatively well-standardised across countries. We first consider simple differentials across space, between cities, towns and rural areas in four regions of the world. Some outcomes, like access to utilities, primarily reflect public service provision differentials. Many health or education outcomes reflect individual demand differentials, potentially influenced by local social and cultural environments, in addition to supply differentials in public and private facilities (e.g., schools, teachers, clinics and health workers). A subset of health outcomes are worse on average in cities, likely due either to their association with wealth and/or stress (obesity, diabetes, high blood pressure) or air pollution (cough, asthma). Finally, attitudes towards violence and crime, use of modern contraception, and fertility may reflect an especially strong influence of peers through the local social environment. Given these different channels, different sets of outcomes are likely to be influenced by very different sets of policy measures including infrastructure investment, training of professionals, education and dissemination of public health information. Other outcomes may have broader policy implications. For example fertility and attitudes towards birth control and ideal number of children are important for thinking about how urbanisation may constrain national population growth, especially in Africa where fertility remains very high.

After looking at cross-sectional differentials in surveys circa 2013, we consider changes over time in a subset of key outcomes that are well-recorded for surveys circa 2000 in addition to 2013. Then we turn to the role of sorting on observables in explaining rural, town, and city differentials. We use a regression approach where we add controls for household and

[1] For a comparison of these, along with GRUMP, see Potere et al. (2009); Potere and Schneider (2007).

individual characteristics to ameliorate sorting. Finally, for Sub-Saharan Africa we explore the role of political status in explaining differentials in outcomes.

2 Related literature

As noted above, urban-rural differences in various outcomes in the developing world have been widely documented. Most critically, Young (2013) finds large differentials in real consumption, including specifically between rural-to-urban migrants and their rural-stayer counterparts. In related work, Chauvin et al. (2017) argue that, while wages rise with city size, in many developing countries cost-of-living does not do so to the same extent. These differences are inconsistent with a very simple model of spatial equilibrium, in which identical people immediately move to the place that offers them the highest real wage.

One potential explanation is that better amenities in rural areas compensate for lower real incomes. A recent literature has investigated in more detail the spatial patterns of not just real income differentials, but also many amenities. In some cases this work, like ours, also highlights issues in defining and classifying urban areas. Ferré et al. (2012) focus on city size gradients in measured consumption poverty. They report that poverty is both deeper and more widespread in small towns than in cities in a set of eight low- and middle-income countries around the world. Corker (2017) compares household services, child mortality, and fertility rates along a similar city size continuum. She finds a similar gradient in these measures, and separately, she finds that small cities that are adjacent to the largest cities (based on GRUMP), though administratively separate, are also relatively advantaged. Dorélien et al. (2013) explicitly compare DHS urban designations with GRUMP (Balk et al. 2006; CIESIN, IFPRI, World Bank, and CIAT, 2011), a dataset that, like GHSL, combines satellite and census data to define urban areas. They find that many areas identified as rural by DHS are peri-urban, with many urban characteristics. We will explore these issues below in our own data. We find that nearly all our measures are better on average in urban than rural areas. However, a few related to urban diets, stress, air pollution, and crime are worse in some or all regions.

Such differentials motivate urban-rural migration, as has been well-documented in some contexts. For one perspective, Brueckner and Lall (2015) review a literature discussing the influence that public services have on rural-urban migrants. In Brazil, although higher wages are the primary driver, the promise of better public services such as water and sanitation also push people into bigger cities, especially the poorest, who may be willing to accept lower wages in exchange for better public amenities (Lall et al., 2009). These ideas are the foundations of the literature on bias toward big and politically powerful cities (Ades and Glaeser, 1995; Davis and Henderson, 2003).

An alternative explanation for these persistent gaps is that the people who live in cities are different than their rural counterparts, in observable ways such as educational attainment, or unobservable ways like intrinsic motivation, that provide them with distinct opportunities. Following Lagakos and Waugh (2013), Young (2013) argues that this kind of selection, or sorting, can largely account for these real income differentials. This is consistent with a large literature on sorting along the city size distribution in the United States, where the share of educated workers increases with city size, but so do the cost of living and the quality of urban amenities (Behrens and Robert-Nicoud, 2015; Bacolod et al., 2009). Kahn and Walsh (2015) discuss endogenous choices made by highly educated workers across cities. They argue that some cities offer better local public goods to attract the highly educated, who in turn outbid the less educated for housing in those cities. Diamond (2016) has an explicit structural model of sorting, city size, worker types and amenities. This idea of sorting motivates our use of regressions to control for individual and household attributes such as age and education when measuring differentials across space.

Another literature has built dynamic structural models with frictions on movement to explain why real income and amenity gaps persist across space. There is work on China focusing on the high costs of moving and birthplace attachment (Tombe and Zhu, 2019), and related work on Indonesia (Bryan and Morten, 2019).

The work most closely related to ours is Gollin et al. (2018). Rather than using city size classes, they consider differentials across population density quintiles or deciles. They find that a wide variety of disamenities, including pollution and crime, decrease with population density, while consumption increases.[2] Given the lower levels of consumption in less dense areas, they interpret this as inconsistent with static models of spatial equilibrium. A more precise interpretation would be that we are in a spatial equilibrium at any point in time, and regional preferences or high moving costs are responsible for observed differences across space in consumer welfare.

Relative to these studies in general, we explore a wider range of countries and a wider range of outcomes, and use the new GHS-SMOD typology differentiating towns and peri-urban areas from larger cities. We also study change over time between 2000 and 2015, with comparable survey data and consistent city and town definitions.[3]

3 Data

In this investigation we rely primarily on two data sources: the degree of urbanisation layer (GHS-SMOD) for the year 2015 from the EU's Global Human Settlement Layer (GHSL), along with the corresponding underlying population data GHS-POP, and Demographic and Health Surveys (DHS). For comparisons over time, we use analogous data circa 2000.

3.1 Global Human Settlement Layer

The Global Human Settlements Layer (GHSL) is a suite of global grids, produced by the European Commission's Joint Research Centre, documenting built-up area, population, and settlement type for four epochs, circa 1975, 1990, 2000, and 2015. It consists primarily of three data products: GHS-BUILT, GHS-POP, and GHS-SMOD.[4] The first, GHS-BUILT (Corbane et al., 2018, 2019), reports a binary measure of whether each approximately 30-by-30-meter pixel on land is built up or not, based on data from the Landsat series of satellites from a year near to that epoch's nominal year. The second, GHS-POP (Schiavina et al., 2019; Freire et al.,

[2] In earlier work, Balk et al. (2004) report a density gradient in childhood mortality.

[3] While the GHS also reports town and city boundaries for 1990 and 1975, neither DHS surveys prior to 1996 (except in Egypt) nor their precursor World Fertility Surveys in the 1970s and early 1980s had the Global Positioning System (GPS) location information to allocate surveyed families accurately over space.

[4] A fourth product reports built-up area for the fourth epoch using higher-resolution data from the SENTINEL satellite.

2016), reports the population of each 250-by-250-meter pixel, based on CIESIN, 2017 and GHS-BUILT. CIESIN, 2017 generally reports populations of the smallest available census units for all countries, allocated to 1 km^2 pixels assuming uniform density within each census unit; the version used here performs the same reallocation for 250-by-250-meter pixels. GHS-POP further reallocates population within census units proportional to the share of each 250-by-250-meter pixel that is built-up according to GHS-BUILT.

Finally, the Degree of Urbanisation (GHS-SMOD L1; Pesaresi et al. 2019) assigns 1-by-1 km pixels to one of three categories based on population density, contiguity, and built-up share: cities, towns/suburbs, and rural areas.[5] Pixels are candidate members of a city if they have a population density over 1500/km^2 or a built-up share over 50%. A contiguous set of such pixels constitute a city if its total population is over 50,000. Towns/suburbs are defined analogously, but at lower thresholds. A candidate town cell must either fulfil one of the two criteria for city candidate cells, or it must have both a population density over 300/km^2 and a built-up share over 3%. A contiguous set of such pixels with a total population over 5,000 constitutes the union of cities and towns/suburbs. After excluding the pixels that belong to a city, the remaining cells are towns/suburbs. Rural cells constitute all the cells that do not belong to a city or towns/suburbs. Further information about the GHSL data is available in Florczyk et al. (2019).[6]

For a variety of reasons one might like to distinguish suburbs from towns. We do not do that here for two reasons. First, by this definition, suburbs are a very small part of the national population in developing countries. Second, suburbs and towns are both poorly represented in the DHS. Given limited samples and the issue of displacement of recorded locations to be discussed below, we were not confident that we could make a meaningful distinction.

[5] The Degree of Urbanisation's second level (L2) further subdivides settlements into a total of seven categories. We do not use this classification because some categories contain few DHS clusters, and their small spatial size makes it likely that cluster displacement generates a large amount of classification error (see below).

[6] Data for India are based on Balk et al. (2019) without the 3% built-up area criterion.

3.2 DHS

With more than 400 surveys carried out in over 90 countries since 1984, the DHS is one of the largest survey programs in the developing world. Beyond a core of questions on birth histories, contraception, and child health and mortality, the surveys have expanded to include a wide variety of questions related to household assets, domestic violence, and other topics. The surveys are carried out by national ministries of health or equivalents, with technical assistance from the United Stated Agency for International Development. While individual countries add questions about topics of particular interest, base survey questions are highly standardised across countries, especially for core topic areas and within each five-year round of the program.

DHS surveys are nationally representative, typically using the most recent census as a sampling frame, and often representative for first-level subnational administrative units (regions, provinces, departments, or states) or aggregations thereof. They generally follow a two stage sampling design, first selecting several hundred census enumeration areas (clusters), and within each of these, selecting 20–30 households. To account for this sampling design, and sometimes also to account for oversampling of certain populations of interest, sampling weights must be applied to generate nationally representative statistics.

A small set of questions is asked of a household representative about all household members, generally defined de facto as those who slept in the household the night before the survey. A much wider set is presented to adults of childbearing age (defined as 15–49 for women and nearly always 15–59 for men) and to mothers or other adult caregivers in the household about children under 5.[7]

Since 2003, DHS has distributed GPS coordinates for surveyed clusters in most, but not all surveys, allowing researchers to link survey data with geographic data not collected in the surveys. In our case, this means linking them to the GHS-SMOD definitions of cities, towns and rural areas. In order to enhance the confidentiality of survey respondents, beginning in

[7] These age bounds apply in all of our sample countries. In some countries, primarily in the Middle East, North Africa and South Asia, the female sample is limited to ever-married women. Bangladesh is the only country in our sample subject to this restriction.

2008 the DHS displaced the reported locations by up to 2 km for DHS-urban clusters and 5 km for DHS-rural, with 1% of clusters further displaced up to 10 km.[8] Displacement direction and distance is drawn from a uniform random distribution, subject to the constraint that the displaced location must be on land, in the same country, and in the same administrative unit (at an administrative level that varies across surveys). Uniform displacement by distance means that the fraction displaced 0–1 km is the same as that displaced 4–5km, despite the large difference in area between those two rings. These changes were applied retroactively to archived data for pre-2008 surveys. In Section 4.1 and Appendix E, we present a methodology we developed to calculate the probability that a household actually lives in a GHS-defined city, town, or rural area conditional on its displaced location.

Relative to GHS-POP estimates, we find that even after applying both DHS sampling weights and using the aforementioned methodology to correct for displacement, DHS over-samples rural area, under-samples in cities, and especially under-samples towns and suburbs. For example, in our sample of Sub-Saharan African countries, rural areas house 27% of the population but 49% of DHS household members; and cities house 45% of the population and 36% of DHS household members. The contrast for towns is even starker. While towns and suburbs house 28% of the population they only account for 16% of DHS household members. Numbers for South Asia are similar.[9] As noted above, this undersampling is one reason why we aggregate towns and suburbs. However, at the level of world regions, overall absolute regional sample sizes are large enough to allow for relatively precise estimates of our indicators of interest.

3.3 Sample of countries

In our main cross-sectional analysis, we limit our attention to “standard” DHS household surveys (as opposed to e.g. interim or health facility surveys) carried out between 2010 and

[8]DHS urban/rural designations are based on national administrative definitions that vary by country. In order to avoid confusion with GHS-SMOD terminology, we refer to them as DHS-urban and DHS-rural.

[9]Prior to correcting for displacement and applying DHS sampling weights, undersampling was much more pronounced. For example, in Sub-Saharan Africa 74% of DHS households were in rural areas, 20% in cities, and only 6% in towns and suburbs.

2016, because GHS-SMOD is benchmarked to 2015 and based on census data from a year close to 2010 in most countries, and thus pre-2010 surveys reflect conditions that might be very different, especially in the context of rapid urbanisation. We further restrict to surveys with available GPS data. This leaves us with data for the following countries:

- Sub-Saharan Africa:
  - Eastern Africa: Burundi, Comoros, Ethiopia*, Kenya*, Malawi*, Mozambique, Rwanda, Tanzania*, Uganda*, Zambia, Zimbabwe*
  - Western Africa: Benin*, Burkina Faso*, Côte d'Ivoire*, Ghana,* Guinea*, Liberia, Mali*, Nigeria*, Senegal*, Sierra Leone, Togo*
  - Central Africa: Angola, Cameroon, Chad, DR Congo, Gabon
  - Southern Africa: Lesotho, Namibia*
- South Asia: Bangladesh*, India, Nepal*
- Southeast Asia: Cambodia, Myanmar, Philippines, Timor-Leste
- Latin America and the Caribbean: Colombia, Dominican Republic, Guatemala, Haiti, Honduras

For the subset of countries with an asterisk in the list above, primarily in Eastern and Western Africa, we also have a comparable survey from 1997–2003, which we combine with GHS-SMOD cities and towns defined circa 2000 to look at changes over time. Appendix Table A shows the full sample of surveys, with their dates and sample sizes.

We focus on results for Africa and South Asia, because their samples are much more representative of their regions. While the South Asia sample is only three countries, one of them is India, home to a large majority of the region's population and 18% of the world's population. By contrast, a very small set of relatively poor countries in Latin America and Southeast Asia are sampled, reflecting the priorities of the DHS program. Results for sampled countries in these regions are unlikely to be representative of each region as a whole.

3.4 Other data sources

For a smaller subset of countries in Sub-Saharan Africa, we extend our 2015 cross-sectional analysis in two ways.

First, we use Afrobarometer data to supplement our analysis on violence. Where DHS asks only questions about attitudes and incidence of domestic violence, Afrobarometer conducts national public opinion surveys including questions on the perception and incidence of crime outside the household. Specifically, we refer to data on fear as well as incidence of both physical attacks and home robberies within the past year. We are able to use round 6 Afrobarometer data from all countries in Sub-Saharan Africa coinciding with DHS sample circa 2015, excluding Angola, Democratic Republic of Congo, Chad, Comoros, Ethiopia, and Rwanda.

Second, we extend our 2015 analysis using DHS data to include the effect of political and administrative status by employing Statoids data on national and regional capitals as controls. Statoids contains a history of administrative divisions throughout the world.[10] We source geo-coordinates of Statoids' list of capitals in Sub-Saharan Africa in 1990 from Google's geocoder API. We assign DHS clusters either no political status, national capital status, or regional capital status depending on whether clusters are located in or near cities. Specifically, we determine whether a national capital point falls within in a polygon defined as a city in GHS-SMOD, or a regional capital point falls within polygons defined as either cities or towns in GHS-SMOD. We then draw a 10km buffer around the city or town polygons in question and assign the corresponding (national or regional) capital status to all clusters within the buffer. We draw a 10km buffer around the remaining regional capital points and assign clusters within them regional capital status. All other clusters are assigned as having no political status. Roughly two-thirds of DHS clusters are in areas with no political status.

[10] http://www.statoids.com/

3.5 Outcomes

While each DHS survey contains hundreds of questions, we generally consider individual and household attributes that are of particular policy interest, that have information available for a large number of countries, and for which definitions are well-standardised across surveys. These variables fall within five broad categories: utilities, health, education, fertility, and violence. For a few health items of interest, our data are limited to India. For more information about these measures and how they were obtained, see Croft et al. (2018).[11]

3.5.1 Utilities

We first consider three core household utilities: electricity, water, and sanitation. While the presence of electricity in the household is comparably well-defined, water supply and sanitation are somewhat more problematic. Unlike other indicators we consider, different surveys offer substantially different possible answers. We address this in two ways. First, we consider a subcategory of safe water that is well-standardised: piped water into the house. The analogue for sanitation, flush toilets connected to a sewer system, is rare even in cities of Africa and South Asia, so we don't report it. Second, we use categories that adhere to definitions used by the WHO/UNICEF Joint Monitoring Programme for Water Supply, Sanitation and Hygiene (JMP) for "safely managed" water and "improved" sanitation. These are harder to define uniquely using the available data, but are more closely aligned with policy targets such as the Sustainable Development Goals (SDGs).

3.5.2 Health

Health outcomes are best defined and reported for children. The infant mortality rate, the share of children born live in the 5 years before the survey that did not survive to age 1, is the best indicator in these respects, because respondents are likely to remember infant deaths,

[11] Except where otherwise indicated, survey question quotes are from the Tanzania 2015–16 DHS Final Report (MoHCDGEC, MoH, NBS, OCGS, and ICF 2016.). Exact wording will differ across surveys, at minimum due to language differences, but DHS places considerable emphasis on comparability, so substantive differences are likely small.

but it reflects only extreme events. Our morbidity indicator is the share of children under five experiencing diarrhoea in the past two weeks. This has the advantage of being directly related to water and sanitation infrastructure that we consider above, but it is likely not recalled as well or defined as consistently across respondents. Our measure of health care access/usage for children is the third dose of the diphtheria-pertussis-tetanus (DPT) vaccine. While all vaccine outcomes are well-standardised, other vaccines including measles and earlier DPT doses have extremely high prevalence in most contexts we study, allowing for little variation across locations.

We also consider health care access for adults by using questions about barriers to health care. Specifically, we calculate the share of women reporting distance and, separately but not mutually exclusively, lack of permission as "a big problem" when they "are sick and want to get medical advice or treatment."

Finally, we look at a set of health outcomes that are potentially worse in cities due to more pollution, less exercise, more stress, and a diet with more sugar, fat, and highly processed grains. These outcomes are cough (past two weeks) for children, obesity (a body mass index, or BMI, above 30) for adults worldwide and diabetes, asthma and high blood pressure for adults in India. Diabetes and asthma are only recorded for India during our period of interest; high blood pressure is recorded for India and Nepal but to be comparable we ignore Nepal. Obesity and high blood pressure are measured by survey staff; the rest are self reported.

3.5.3 Education

Our primary education measure is the share of 16-year-olds who have completed at least 8 years of education. We further consider the share of children at ages 12, 13, 14, 15, and 16 with at least 4 years of schooling completed, to understand the extent to which children catch up when they fall behind the typical progression pattern, as well as the share of children age 13 in school. Catch-up is an especially strong feature in the data in Sub-Saharan Africa, where a substantial share of children age 13 and 14 attend grades 3 or 4.[12]

[12] DHS surveys report educational attainment based on an individual's years completed at their highest level (primary, secondary, or tertiary), along with country-specific definitions of the length of each level, presumably

3.5.4 Fertility

Our core fertility measure is the total fertility rate (TFR). This is a synthetic cohort measure that sums the recent (past three years) fertility experience of women across five-year age groups (15–19, 20–24,...,45–49) to measure the total number of children born to a hypothetical women experiencing current period age-specific fertility levels throughout her childbearing years. We also consider respondents' ideal number of children, separately for men and women, and relate it to the TFR.

In addition, we consider, separately for men and non-pregnant women, and separately for all and the unmarried, the share of sexually active (within the last 4 weeks) respondents aged 20–40 using modern contraception, defined as "male and female sterilisation, injectables, intrauterine contraceptive devices (IUCDs), contraceptive pills, implants, male condoms, emergency contraception, and lactational amenorrhea method (LAM)". Because these sample groups include people who actively want children, to focus on family planning we further consider the subset of women who report they do not want to have a child within the next 2 years and also the subset of unmarried males.

3.5.5 Domestic violence and crime

Our last core category is domestic violence and crime. Our key measure of domestic violence is the rate of actual victimisation of currently married women aged 15–49 by their spouse. Specifically, women are asked if this spouse has ever "Forced her with threats or in any other way to perform sexual acts she did not want to; Physically forced her to perform any other sexual acts she did not want to; Physically forced her to have sexual intercourse with him when she did not want to; Threatened her or attacked her with a knife, gun, or other weapon; Tried to choke her or burn her on purpose; Kicked her, dragged her, or beat her up; Punched her with his fist or with something that could hurt her; Twisted her arm or pulled her hair; Slapped her; Pushed her, shook her, or threw something at her". We also report attitudes

to minimise counting grade repetition as attainment. To the extent that these country-specific definitions are subject to error, it is unlikely to affect the 4-year rates, as primary school is rarely less than four years.

toward domestic violence, specifically rates of men and women aged 15–49 who believe wife beating is justified for any reason.

We complement the results on domestic violence with an analysis of crime external to the household using Afrobarometer data circa 2015. For a subset of countries from the Sub-Saharan Africa DHS sample, we report actual incidence of two types of crime (home robberies and physical attacks), and two measures of household attitudes towards crime (ever feared crime in one's own home, or felt unsafe walking in one's neighbourhood).[13]

4 Methods

4.1 Assigning clusters to city/town/rural zones

Using GIS software to assign DHS clusters, and therefore the people within them, to a city, town or rural GHS settlement type is straightforward. However, as noted above, the reported DHS cluster locations are displaced up to 5 km from their actual location. This is especially problematic given that some settlements, especially towns, are quite small in spatial extent. This could possibly be part of the reason why the share of DHS respondents falling within towns is so much smaller than the population share of towns according to GHS-POP.

While we cannot undo this displacement, we can use its structure to improve our estimates. Specifically, rather than assigning a respondent in a cluster uniquely to a zone (city, town or rural), we assign them probabilities for each zone. We draw a disc centered around each reported cluster location with a radius corresponding to the maximum displacement of its DHS type (2 km for DHS-urban, 5 km for DHS-rural). The probability that a cluster's true location lies in each of the three zones is a function of the overlap between the zone and the disc. It is not simply the ratio of the overlapping area to the area of the disc for two reasons. First, the displacement probability distribution is not areally uniform throughout the disc; the displacement distance distribution is uniform, which means that points are more

[13] We lack Afrobarometer data for the following African countries in our DHS c. 2015 sample: Angola, Democratic Republic of Congo, Chad, Comoros, Ethiopia, and Rwanda.

likely to be in the centre of the disc. Second, since sampling is proportional to population, more populated places are more likely to be the true location. To take an extreme example, an unpopulated place cannot be the true location. In practice the probability P_{zj} that the true zone type of cluster j is $z \in \{\text{city}, \text{town}, \text{rural}\}$ can be well approximated by dividing the disc into concentric rings of width 0.2 km and calculating:

$$P_{zj} = \frac{\sum_s A_{zjs}/r_s}{\sum_k \sum_s A_{kjs}/r_s} \tag{1}$$

Where r_s is the central radius of ring s, and A_{zjs} is the population in zone z in cluster j in ring s. We use the GHS population grid GHS-POP (at 250-meter resolution) as our measure of population. In calculating the indicators described above for e.g. town residents, we treat a person with a probability P of living in a town as a fraction P of a person. This is analogous to how we apply the sampling weights below. For more information, see Appendix E.

To validate this method, we use data from five surveys from 1998–2003 (Burkina Faso, Ghana, Mali, Niger, and Togo) for which we have both the original reported locations and their displaced analogues. Consistent with the prediction above, 73% of clusters whose true location is a town are displaced to other zones; the analogous rates for rural and city clusters are 5% and 29%. We also ran regressions on this crosswalk sample to validate our method. In many cases, displaced, true and estimated location types yield very similar results. In some cases, both estimated and displaced yield similar results which noticeably differ from those for true locations. However in another set of cases, using estimated locations does yield better results than displaced, in the sense that they are closer to those based on true locations. In summary, using our estimated locations either improves upon or leaves unchanged results from using displaced locations for this sample of 5 surveys.

4.2 Sampling weights

As noted above, in all analysis we weight observations. First DHS surveys require sampling weights to be representative at the country level and we always apply these weights. Second, since our analysis is mostly at the level of supranational regions, in order to produce estimates that are representative of these regions we must adjust the weights to reflect different country sizes. For example, while fewer households are sampled in Nigeria than in Mali, Nigeria represents a much larger fraction of Sub-Saharan Africa's population. For the basic cross-section and over-time graphical analysis, for each location type in a country, we calculate an average outcome such as the total fertility rate, across clusters (using DHS sampling weights for each cluster as well as adjusting for the estimated probabilities of each cluster being in a specific location type). We then calculate a weighted average, across countries within the supranational region, of these country cluster averages for each location type, where the weights are each country's share of the supranational region's GHS population of the corresponding location type. For instance, we calculate the total fertility rate (TFR) for location type i in a region of countries indexed by j as follows:

$$\mathrm{TFR}_i = \frac{\sum_j \mathrm{TFR}_{ij} \cdot p_{ij}}{\sum_j p_{ij}} \tag{2}$$

where p_{ij} refers to the GHS population of location type i in country j.

4.3 Regression methods

In addition to reporting average values for each indicator for each location type, we apply regression analysis to adjust for differences in age, sex, and other attributes that may be the source of differences in raw outcomes. For example, if rural areas have a higher infant mortality rate than urban areas, this could be due to the fact that mothers give birth at earlier ages or that mothers are more poorly educated. Regression analysis allows us to measure the urban-rural difference having removed (under certain assumptions) the influence of age and education differences. It also allows us to quantify uncertainty due to finite samples (though not uncertainty due to measurement error in either outcomes or settlement boundaries) by calculating standard errors.

The general regression equation we estimate is:

$$y_{ijc} = C_{jc}\beta_0 + T_{jc}\beta_1 + \alpha_c + X_{ijc}\gamma + \epsilon_{ijc} \tag{3}$$

where y_{ijc} is an outcome for respondent i in survey cluster j of country c, C is a probability of being in a GHS-SMOD city, T is a probability of being in a GHS-SMOD town, and thus rural is the reference category. Our coefficients of interest are β_0 and β_1, the differentials from living in a city or town relative to a rural area. α_c is a survey fixed effect, equivalent to a country fixed effect in the main cross-sectional specification, and X is a vector of control variables. We have four types of survey subjects: children (under 5), teens, adults (separately by sex), and households. For all of them, we control for the sex, log age, and education of the household head, where education is entered with dummy variables for primary, secondary, and post-secondary (relative to a base of no primary). For teen regressions (on schooling), we add a control for sex of the teen.[14] For adult female outcomes, we control for a cubic in age and the education dummy variables. For children, we control for child sex, log child age, a cubic in age of the mother at child's birth, and education dummies for the mother. Standard errors are clustered by survey cluster j.

When analysing unadjusted averages in the raw data, our outcomes of interests are rates for a population of interest. In the regression analysis, since our unit of analysis is the individual or household, the outcome variables are either counts or indicators for whether a given outcome is true for the individual in question. In the case of fertility, for example, rather than a total fertility rate, the outcome is a woman's number of births in last 3 years. For infant mortality, the analogous outcome is whether each child born in the last five years died before 1 year of age. To avoid censoring problems, we restrict the sample to children born 1 year or more before the survey. Other outcomes like utilities or having vaccinations are defined the same in the raw data and regressions, but of course in regressions they are represented as individual outcomes, rather than regional averages.

Finally the observations in regressions are weighted. As always the DHS sampling weights are applied. These are further scaled to account for the fact that sample sizes across countries are not proportional to their national populations. Overall, the regression observations are thus weighted by:

$$W_{ij} = w_{ij} \frac{p_j}{\sum_{i \in j} w_{ij}} \tag{4}$$

[14]Unfortunately we cannot assign teens to their parents consistently to derive their parents' attributes.

where w_ij is the DHS sampling weight for respondent i in country j, and p_j is the GHS population of country j.[15] In essence, we rescale the DHS sampling weight by the ratio of population size to sample size of its country.

5 Results

5.1 Baseline: Graphs for circa 2015

In most cases, we see better outcomes in cities than in rural areas. Towns and suburbs often (but not always) fall somewhere in between cities and rural areas, a pattern that we refer to as expected, given, for example, economies of scale and an intermediate position in migration paths. However in our South Asia sample (India, Bangladesh, and Nepal), towns and suburbs often, even typically, have worse average outcomes than rural areas. This could be because some areas characterised by the GHS degree of urbanisation model as town or suburbs in this region, though dense, are highly agricultural, and potentially isolated from features more commonly associated with towns. However it could also be that towns and suburbs have been unable to provide services in the presence of rapid growth. This pattern is so common that when we see towns and suburbs performing worst in South Asia, but in the intermediate position in the other regions, we refer to this as the usual pattern.

5.1.1 Utilities

Figures 1a and 1b show consistently better outcomes for electricity and piped water into the house in cities than in rural areas in all regions. The electrification rate differentials are especially large in Sub-Saharan Africa and Southeast Asia, which also have the lowest average rates. Rates for towns and suburbs are between those for cities and rural areas, except in South Asia which follows the usual pattern. There, electrification rates are almost 10 percentage points lower in towns and suburbs than in rural areas. As noted, piped water is more

[15] For variables on physical spousal violence, the DHS domestic violence sampling weight is used instead of the standard DHS sampling weight.

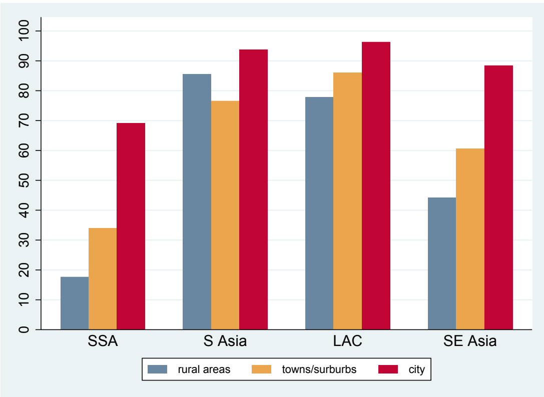
available in cities everywhere; but overall rates in Africa and Southeast Asia are very low, with little difference between rural areas and towns. Latin America, where there is much higher general uptake of piped water, follows more the expected pattern where cities have the most piped water, but towns have a sizeable advantage over rural areas.

Given the low prevalence of piped water and the very low rates of flush toilet into a local sewer system in our two main regions, we also consider broader categories: improved sanitation and safely managed water, which includes protected wells or springs, boreholes, packaged water, and rainwater. Figure 1c shows much higher penetration of safely managed water than piped water alone in all regions, and a more consistent expected gradient, with towns and suburbs at least 5 percentage points higher than rural areas and lower than cities (even in South Asia). Improved sanitation (Figure 1d) also follows the expected gradient, with large city-rural gradients of at least 20 percentage points, and town/suburb levels above but closer to rural areas in most regions. For Sub-Saharan Africa specifically, overall rates are lower than in other regions, and there are very large rural-city differentials. In particular, 40% and 70% of city households have respectively safe water or improved sanitation, while in rural areas the respective numbers are about 12% and 36%, roughly half of city rates.

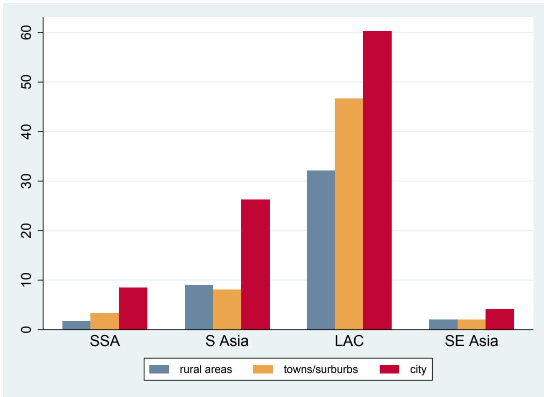
5.1.2 Health

In Figures 2a and 2b, infant mortality and diarrhoea rates follow the usual gradient, but differentials across rural, town and city are generally not large. Urban-rural infant mortality differentials are in the range of 5–10 deaths per thousand live births, for overall rates of about 25–50 by region, with the high end being Sub-Saharan Africa. Diarrhoea differentials are proportionally even smaller, never exceeding a few percentage points. In Figure 2c, DPT vaccination differentials are similarly small, perhaps due to large international vaccination efforts, and in the expected directions, except in Latin America, where towns and suburbs have the lowest rates (but barely so), and cities and rural areas are essentially indistinguishable. These are health dimensions where there are inter-regional differentials, but within-region differences between rural, town and city are not large. When we control for sorting on observables in regressions in section 5.4, these differentials typically get smaller, go away, or

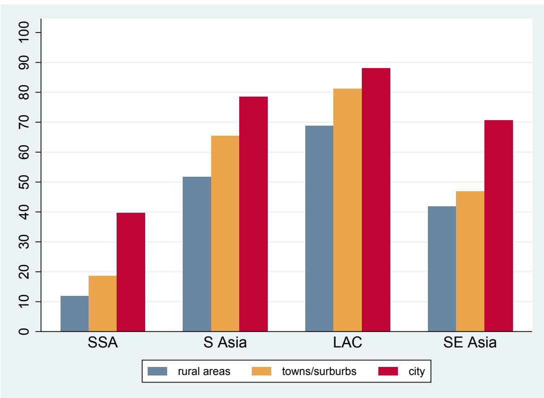
Figure 1: Percentage of households with utilities



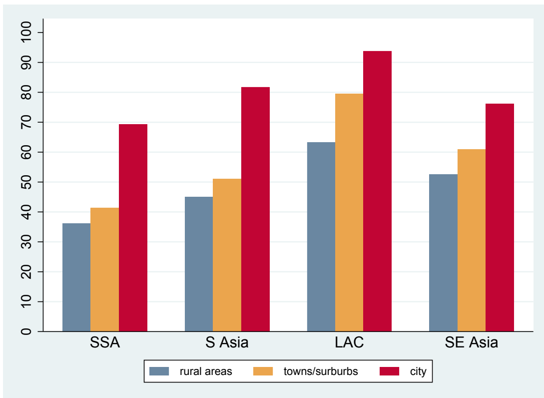
(a) Electricity



(b) Piped water into dwelling



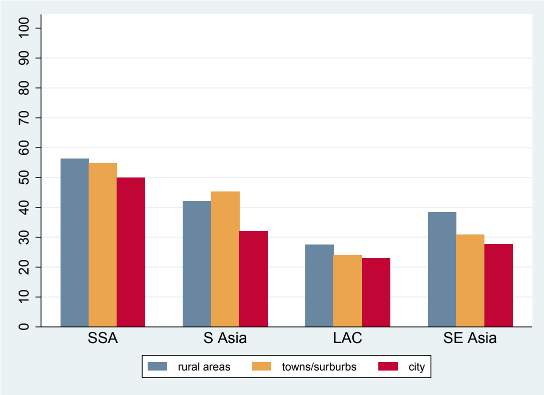
(c) Safely managed drinking water



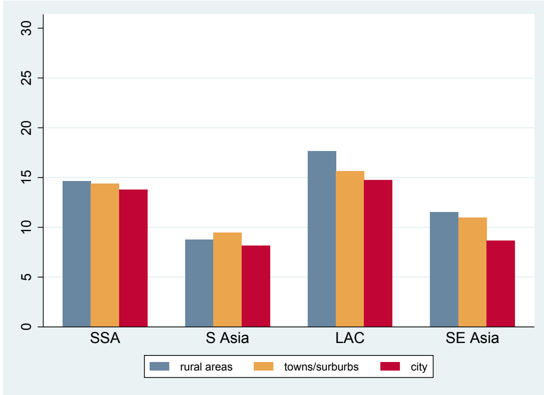
(d) Improved sanitation

Compared to our base set of countries, in panel (a), survey responses for electricity were unavailable for Honduras 2011–12. In panels (b) and (c), safely managed drinking water and piped water into dwelling were unavailable for Cambodia 2014. Safely managed drinking water is defined by the DHS-WHO Joint Monitoring Program as all improved water sources that take zero minutes to collect, or are on the premises. Improved water sources encompass all piped water and packaged water, as well as protected wells or springs, boreholes, and rainwater. Improved sanitation is defined by the DHS-WHO Joint Monitory Program to include: all shared and non-shared facilities that flush/pour flush to piped sewer system, septic tank, pit latrine; ventilated improved pit latrine, pit latrine with slab, and composting toilet. Additionally, facilities that flush to unknown locations are considered improved, whereas facilities that flush to a known location but not to a sewer system, septic tank, or pit latrine are classified as unimproved.

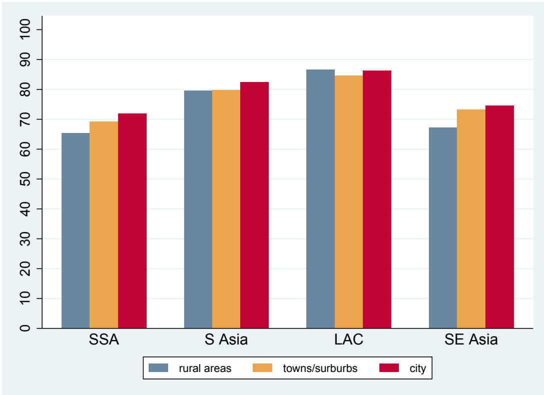
Figure 2: Core child health outcomes



(a) Infant mortality rate (per 1000 live births)



(b) Diarrhoea



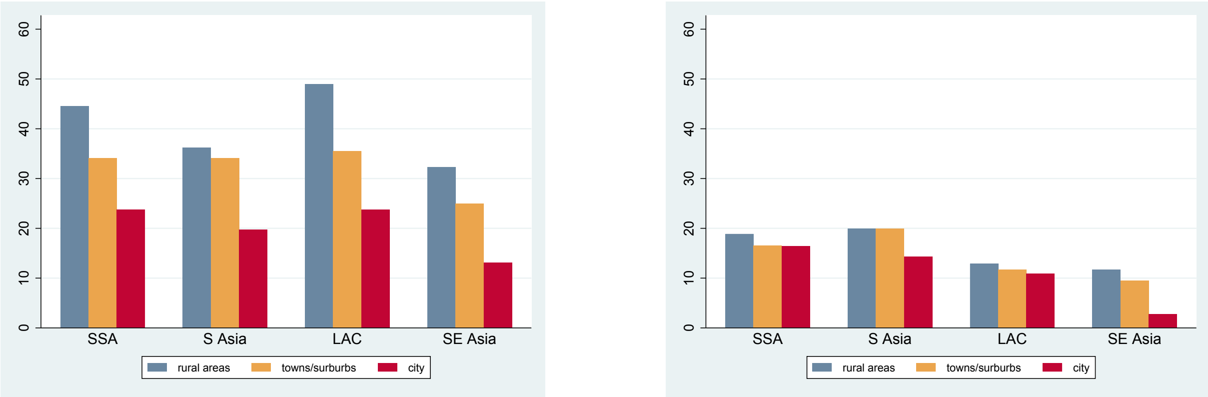
(c) Third DPT vaccination

In panel (a), the infant mortality rate is defined as the probability of dying before the first birthday, per 1000 live births. Panel (b) indicates the percent of children aged 5 and under with diarrhoea in the last two weeks. Panel (c) presents the percentage of children aged 12–24 months that received the third dose of the DPT vaccine.

are even reversed.

Access to health facilities varies much more across location types and in the expected direction in Figure 3a. Not surprisingly, the share who report they do not seek care at a medical facility due to distance is much lower in cities, where facilities are more readily available. This is consistent with cost minimisation in care provision, as fixed costs reduce per capita costs in dense places; but it may also reflect an urban bias in provision. Rural respondents report distance as a reason for not going to a health facility twice as often as urbanites in most regions. In Figure 3b, lack of permission limits women from accessing health care but at much lower reported levels compared to distance, with small within-region differentials in general. The small differentials tend to follow the expected pattern across location types, but

Figure 3: Percentage of respondents who didn't go to a medical facility when sick



(a) Distance (b) Permission

The sample consists of women aged 15–49. We record the fraction of respondents reporting that distance or permission was a big problem versus no problem or a small problem. Data was missing for the following surveys: Colombia 2010, Bangladesh 2014, and Guinea 2012.

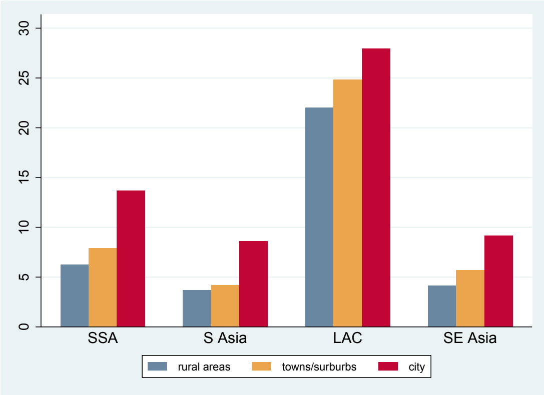
towns and suburbs are indistinguishable from cities in Africa, and from rural areas in South Asia.

Finally, in Figure 4 we consider a set of health outcomes especially likely to be worse in cities, as they are influenced by air pollution, lack of exercise, and the abundance of unhealthy food often associated with cities. Obesity shows striking differentials, with rates in Sub-Saharan African and South Asian cities being 2.3 times those in rural areas, and having monotonic increases with density in all regions. Cities also see slightly higher rates of children's cough than rural areas in all regions. In regressions in section 5.4, even these modest differences for Sub-Saharan Africa and South Asia are statistically significant. We report the remaining three indicators only for India. There high blood pressure, asthma, and diabetes among adults are all higher in cities than in towns, which in turn have higher rates than rural areas. The diabetes rate for cities is twice that of rural areas. These indicators thus provide some evidence that there are costs to urban living that may at least partially offset the many benefits noted here and elsewhere.

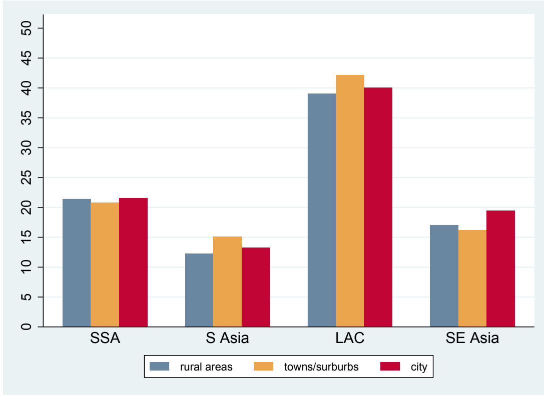
5.1.3 Education

The share of 16-year-olds with at least 8 years of education (roughly lower secondary school completion) follows the usual pattern, although the spatial differentials in South Asia are small

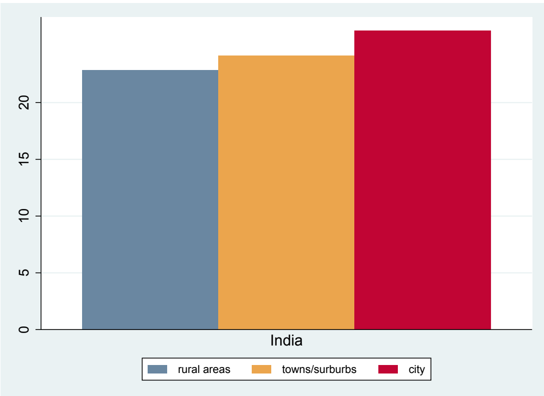
Figure 4: Other health outcomes



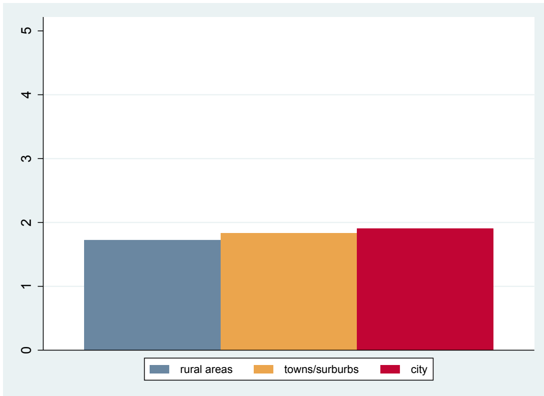
(a) Obesity



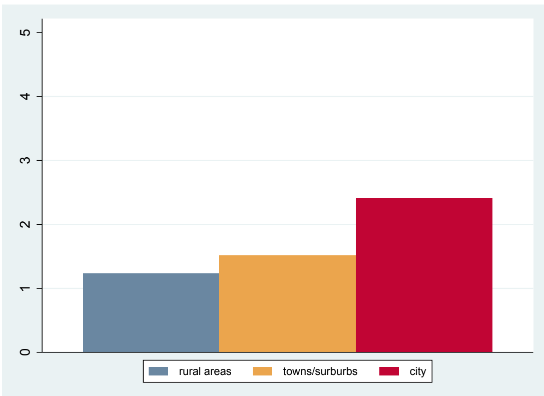
(b) Cough



(c) High blood pressure, India 2015–16



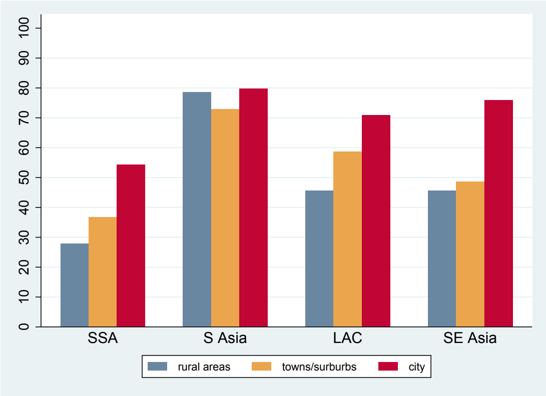
(d) Asthma, India 2015–16



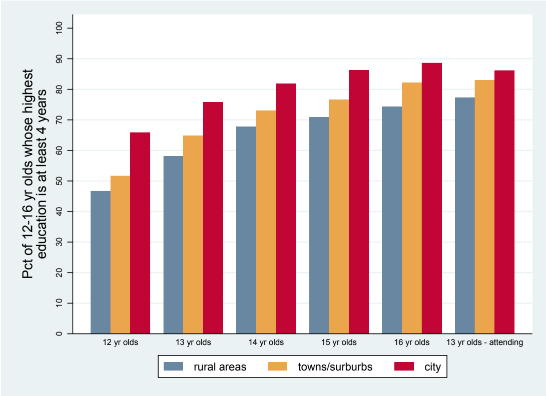
(e) Diabetes, India 2015–16

Panel (a) displays the percentage of obese respondents, where obesity was defined as having a body mass index ≥ 30. The sample consists of household members that were 20–49 years old, eligible for the individual male and female surveys, and not pregnant. Data on obesity for the Angola 2015–16 survey was unavailable. Panel (b) shows the percentage of children aged 5 and under who have had a cough in the last two weeks. Panel (c) shows the percentage of household members aged 25 and over with high blood pressure. High blood pressure is defined as a mean systolic measure ≥ 140 or a mean diastolic measure ≥ 30 over three blood pressure readings. Respondents who reported taking medication for hypertension, or who reported that a medical professional told them they were hypertensive, were also coded as having high blood pressure. Panels (d) and (e) display the percentage of 15–49-year-olds that self-reported asthma and diabetes, for household members eligible for the individual male and female surveys. In panel (d) for diabetes, pregnant women are additionally excluded from the sample.

Figure 5: Schooling



(a) Percentage of 16-year-olds with at least 8 years of education



(b) Education by age in sub-Saharan Africa

In panel (b), education by age is calculated as the percentage of 12–16-year-olds whose highest level of education (in single years) is at least 4 years of schooling. We also calculate the percentage of 13-year-olds who are currently attending or who have attended school in the past year.

(Figure 5a). Differentials are much larger in Sub-Saharan Africa, where rural rates are below 30% and cities over 50%. There are also large differentials for Latin America and Southeast Asia. Cities really do seem to offer much better educational opportunities except in South Asia. Figure 5b offers further information about patterns in Sub-Saharan Africa. The 8-year completion rate by 16-year-old children is reduced by a substantial share of 12-, 13-, and 14-year-old children still finishing their first 4 years of primary school. This suggests two things. Earlier years in school were marked by lack of completion of the grade, and potentially by low attendance. However, on the bright side, rather than dropping out, these children persisted in attending school to at least complete 4 years, which may imply substantial literacy. For example, only 58% of rural 13-year-olds have completed 4 years, but 77% of them are still attending school so that by age 16 the completion rate for 4 years has risen steadily to almost 75%. This pattern is much less pronounced in South Asia, where 13-year-olds who have not finished 4 years of primary school are unlikely to have done so by age 16. Given that, we do not show the corresponding graph.

5.1.4 Fertility

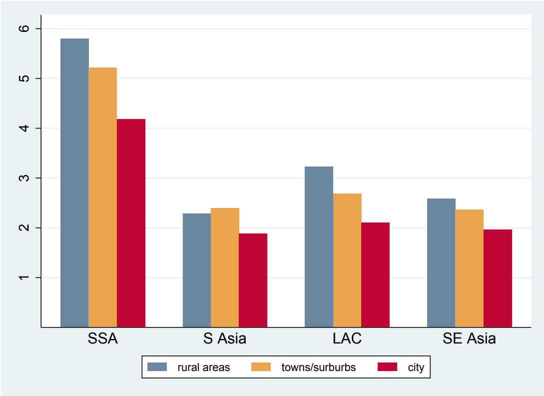
Fertility rates also follow the usual pattern, with especially large differentials between cities and rural areas in Africa and Latin America, where rural rates are almost 50% higher (Figure 6a). These fertility patterns are reasonably consistent with what people report as their ideal number of children (Figure 6b), although in many cases actual fertility is higher then ideal. The spatial differentials in ideal number of children are less than the differentials in actual fertility outcomes; and men generally want more children than women, consistent with the fact that women bear more of the cost.

High fertility in Sub-Saharan Africa is of great concern for the future development of that region. Interestingly, in cities of Sub-Saharan Africa, achieved fertility is actually lower than desired fertility for both men and women. Urban women have over 1.5 fewer children than rural, but they only desire less than 1 more. This is consistent with substantial incentives to not have children in urban areas, such as the higher opportunity cost of women's time to the household and the desire to better educate children as seen in Figure 5a. In other regions, at least for women in cities, the actual is also less than the ideal number of children.

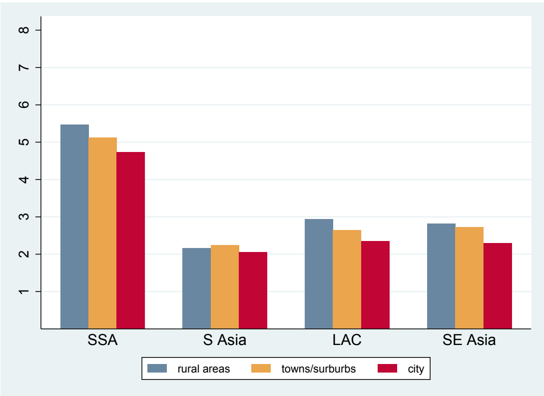
Next we turn to various dimensions of contraceptive use. Contraceptive use tends to mirror fertility, with less overall use in high fertility regions, but there are nuances. Among sexually active non-pregnant females in Figure 7a and among sexually active males in Figure 7b, the spatial pattern across regions follows that for desired number of children for males and females. However now the male-female differentials are even larger. In South Asia in particular, over 55% of women but only about 20% of men report using modern contraception. That begs the question of what men know about their partners' contraception use. Within regions, contraceptive use mostly follows the usual pattern, although urban-rural differentials are generally small. Rural rates are actually modestly higher than urban for men and women in South Asia, and essentially the same for men in Southeast Asia (Figure 7a).

Since contraception is related to the desire to have children, we also look at two subsamples. First, in Figure 7c, sexually active non-pregnant women who report that they do not want to have a child in at least the next two years have somewhat higher use of modern contraception

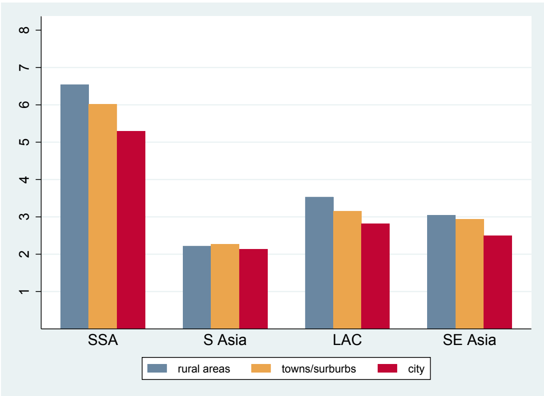
Figure 6: Fertility



(a) Total fertility rate, per woman



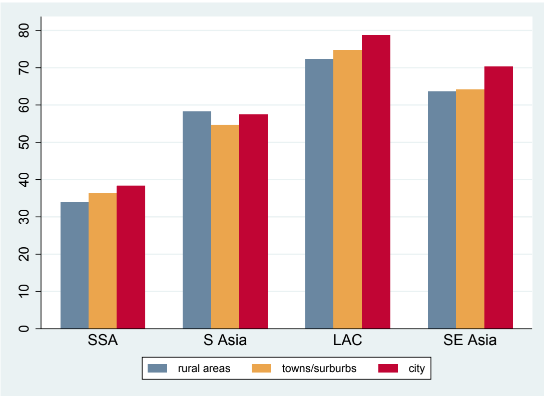
(b) Female average ideal number of children



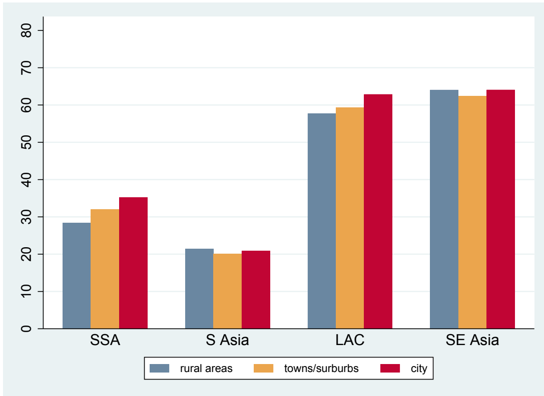
(c) Male average ideal number of children

In panel (a), the total fertility rate is calculated as the number of births per thousand women if they were to give birth at the current (past 3 years) age-specific (15–19, 20–24, ...45–49) fertility rates. In panel (b), the average ideal number of children is calculated for women aged 15–49. In panel (c), the average ideal number of children is calculated for men aged 15–59. In panel (c), data is missing for the following surveys: Colombia 2010, Bangladesh 2014, and Dominican Republic 2013.

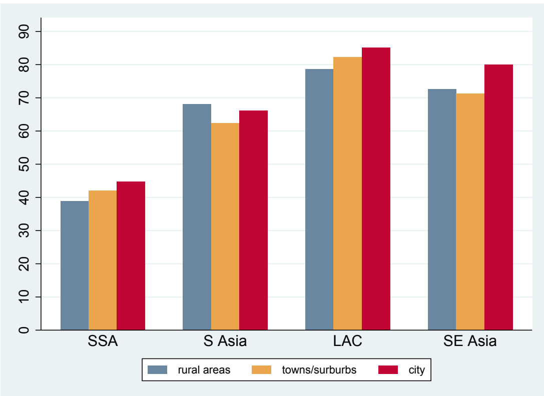
Figure 7: Percentage of respondents currently using modern contraception



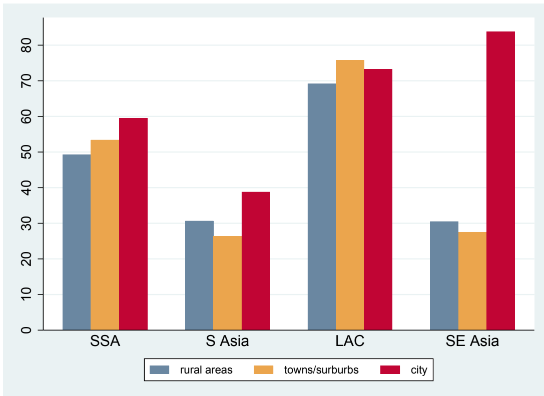
(a) Female



(b) Male



(c) Married females who do not want more children for at least 2 years



(d) Unmarried males

Modern contraception includes the following forms: pill, IUD, injections, diaphragm, condom, female sterilisation, male sterilisation, implants, female condom, foam and jelly, emergency contraception, other modern. The sample consists of men and women sexually active in the last 4 weeks, between ages 20–40. Pregnant women were excluded from the sample. Data on modern contraception was missing for Bangladesh. Data from the following surveys were unavailable, or dropped to make samples comparable across sexes: Bangladesh 2014, Colombia 2010, Dominican Republic 2013, Haiti 2012, and Guinea 2012.

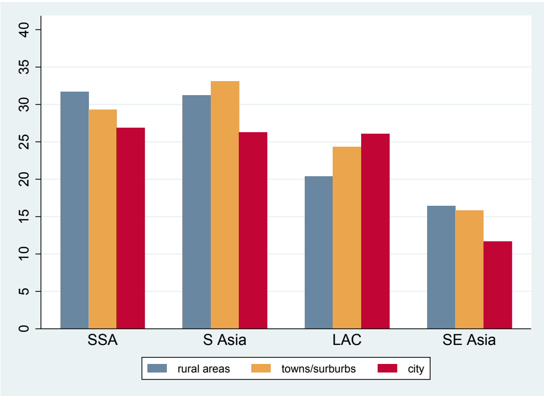
in all contexts, but the spatial patterns are essentially the same as for all sexually active women. Note in Sub-Saharan Africa, this group of women who do not want to have a children any time soon still has absolute low usage of modern contraception, even in cities where the rate is only about 45%. In South Asian cities, the rate is only just above 55%. The second sub-sample is unmarried males, as shown in Figure 7d. Their contraceptive use is higher than for married males in all regions, except Southeast Asia. In Sub-Saharan African cities unmarried males use contraception at almost double the rate of all males, implying extremely low rates for married men. Within-region patterns are usual for Sub-Saharan Africa and South Asia, but Southeast Asia shows very strange patterns. There is very high contraceptive use in cities among unmarried males compared to all males, but the reverse is true to an even larger extent in towns and rural areas.

5.1.5 Violence and crime

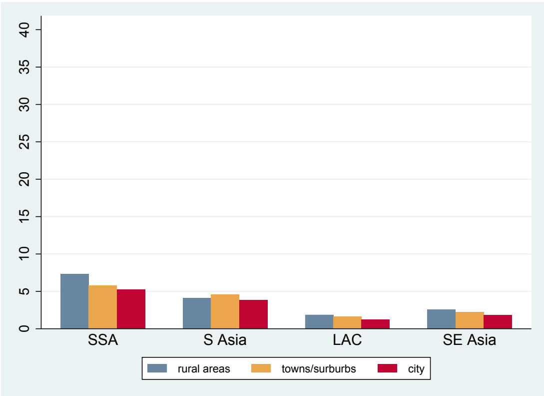
In Figures 8 and 9, we examine attitudes towards and experience of violence and crime. Figure 8a shows that roughly 15–30% of married women in all zones of all regions have ever experienced domestic violence. The shares that report often experiencing violence in Figure 8b are substantially lower, mostly under 5%. Although Southeast Asia has the highest rates of men and women who believe beatings can be justified, they have the lowest actual rates. And relatively more progressive attitudes in Latin America about the justification for domestic violence (Figures 8c and 8d) translate into the best outcomes in Figure 8a, but middling outcomes in Figure 8b. Within regions in Sub-Saharan Africa, South Asia and Southeast Asia, outcomes follow the usual spatial pattern even though we might expect rural women to report violence less. However, the differentials are not large. Latin America again is very different: women in cities and towns report ever having experienced violence more than women in rural areas. Attitudes toward domestic violence also mostly follow the expected pattern, with fewer urban men and women believing that beatings are ever justified, thought none of the differentials are large. Only among men in Southeast Asia are urbanites more accepting of domestic violence than their rural counterparts. Interestingly, substantially more women than men believe that wife-beating is justified for at least one reason in all zones of Africa and

Figure 8: Domestic Violence

Experienced domestic violence

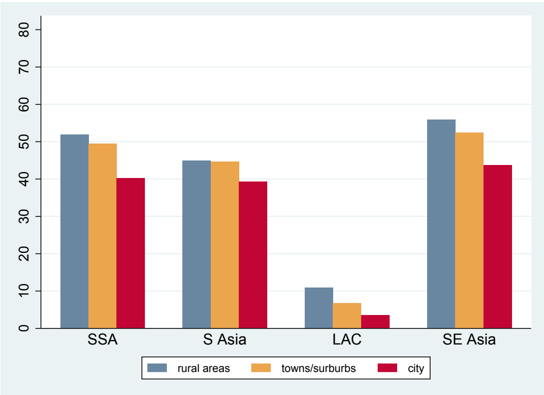


(a) Percentage of currently married female respondents who have ever experienced physical spousal violence

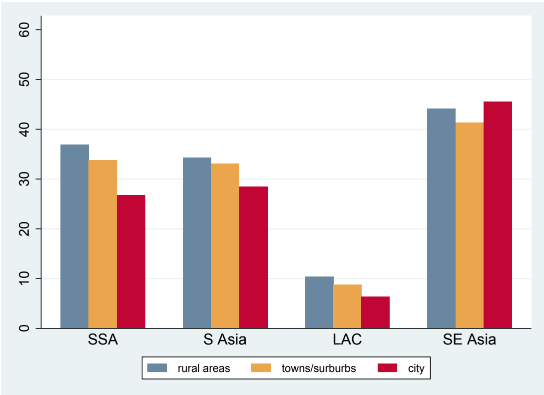


(b) Percentage of currently married female respondents who have often experienced physical spousal violence

Attitudes towards domestic violence



(c) Percentage of female respondents who believed wife beating was justified for at least one reason

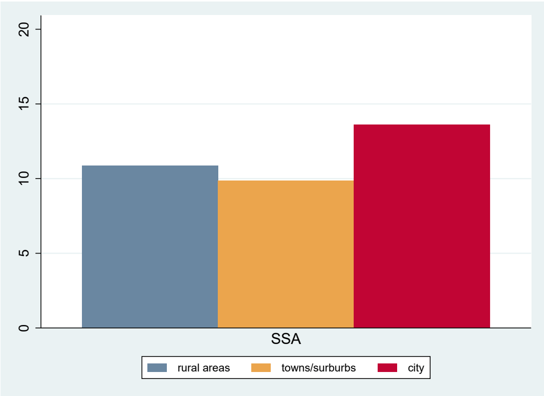


(d) Percentage of male respondents who believed wife beating was justified for at least one reason

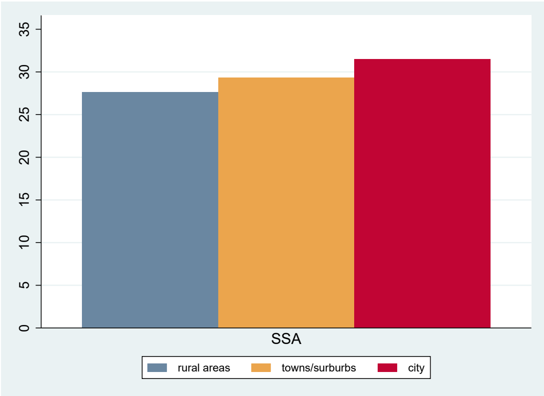
The sample consists of men and women aged 15–49. Panels (c) and (d) further restrict the sample to currently married women. Panels (a) and (b) plot the percentage of respondents whose opinion is that a husband is justified in hitting or beating his wife when: she goes out without telling him, she argues with him, she refuses to have sex with him, or she burns the food. Data on domestic violence in Panels (c) and (d) was unavailable for the following surveys: Bangladesh 2014, Burundi 2010, Lesotho 2014, Benin 2011–12, Ghana 2014, Guinea 2012, Liberia 2013, Senegal 2010–11

Figure 9: External violence

Actual crime

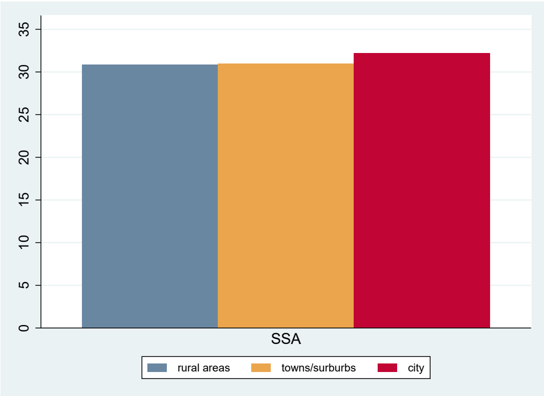


(a) Percentage of households with family members that have been physically attacked in the past year

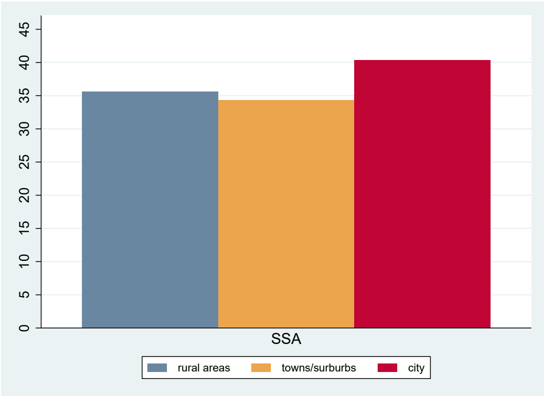


(b) Percentage of households that had something stolen from their house in the past year

Fear of crime



(c) Percentage of households with family members that have feared crime in their own home in the past year



(d) Percentage of households with family members that have felt unsafe walking their neighbourhood in the past year

Data taken from Afrobarometer wave 6 circa 2015. Sample is the African DHS countries minus Angola, Democratic Republic of Congo, Chad, Comoros, Ethiopia, and Rwanda.

South Asia, and in rural areas and towns in Southeast Asia.

In Figure 9, we employ Afrobarometer data for a smaller subset of countries in Sub-Saharan Africa to examine incidence and attitudes towards crime. The share of households with family members who experienced physical violence in the past year is between about 10 and 15% in all zones, and property crime incidence is roughly twice as prevalent. Unlike with domestic violence however, indicators for both actual overall crime and fear of crime are higher in cities than in rural areas. Figures 9a and 9b report roughly 3–4% higher incidence of physical attacks and home robberies in cities relative to rural areas. Further, though differences in fear of crime across space are much less pronounced, Figures 9c and 9d also show greater fear of crime in their home or neighbourhood in cities than in rural areas or towns/suburbs. Overall, we find that different types of violence and crime follow different spatial patterns in Sub-Saharan Africa: urban areas have less domestic violence but more violence and crime outside the household.

5.2 Heterogeneity Across Countries c. 2015

In Appendix B, we further show outcomes circa 2015 for each spatial class by country. Here we plot rural, town, and city coefficients and error bands, effectively showing country averages. For each outcome and graph, we organise countries by ordering rural coefficients from high to low.

In Figures B.1–B.3 in the Appendix, we see common patterns in utilities in virtually all countries. Rural areas are generally poor in electricity, safe drinking water, and sanitation. There are generally large and statistically significant gaps in utilities between all three classifications of rural, town/suburb, and city. Similarly, we see clear patterns in Figures B.9 and B.10, where schooling and fertility show large gaps between cities and rural areas, though towns are often statistically indistinguishable from cities, rural areas, or both: small samples start to matter for precision at the level of many individual countries. Point estimates remain generally higher in urban areas for modern contraception use, though statistical significance disappears in a few countries (Figure B.11).

In Figures B.4–B.8, many countries do not show significant differences for point estimates across space for health outcomes. Obesity is an exception, where there is a clear gap between higher obesity incidence in cities versus lower rates in rural areas.[16] Otherwise, differences in point estimates for infant mortality are small and insignificant across rural areas, towns, and cities. For diarrhoea (B.5), results vary considerably across countries. For cough (Figure B.8), point estimates are generally higher in cities than in rural areas, though differences are not significant. Finally a clearer spatial pattern emerges for with the third dose of the DPT vaccination, where a statistically significant gap between city and rural vaccination rates emerges in Sub-Saharan Africa among countries with the lowest overall rates.

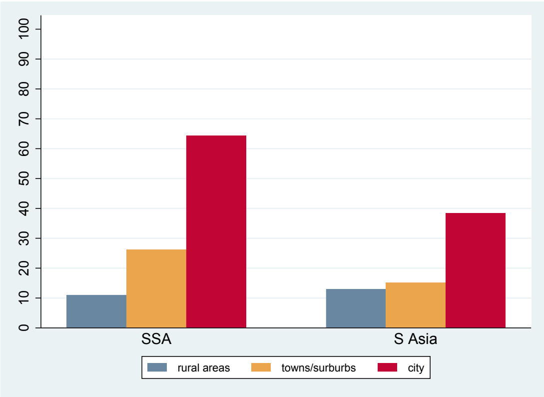
Lastly, we turn to violence and crime in Figures B.12-B.15. Attitudes toward domestic violence show a generally large and significant city-rural gap in most countries, with more progressive attitudes in cities (Figure B.12). Actual victimisation rates show differentials that are mostly but not always in the same direction. They are also smaller and rarely significantly different from zero (Figure B.13). In contrast to domestic violence, urban and rural rates of our other crime measures are generally statistically indistinguishable within countries (Figures B.14 and B.15).

5.3 Changes circa 2000–2015

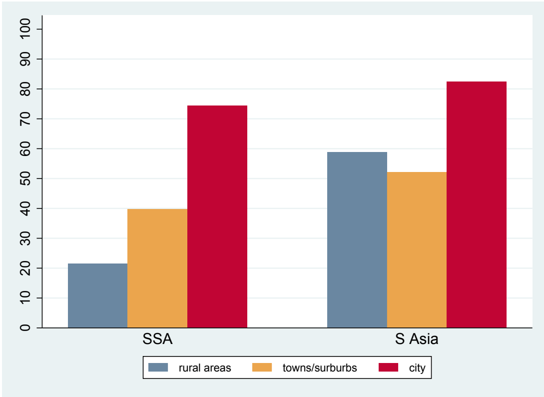
We now turn to changes over time, comparing outcomes circa 2000 (plus or minus 3 years) with those circa 2013 (plus or minus 3 years), although we label the latter 2015 both because many surveys are near that date and because that is the nominal date of the GHS data. Most of the measures analysed above, except for safe water, improved sanitation, reasons for not attending a health clinic, and reported violence, were also available for a survey circa 2000 in much of Sub-Saharan Africa and in Bangladesh and Nepal. Since earlier Indian DHS were not geo-referenced, we have no data on high blood pressure, asthma, or diabetes. Our countries in Latin America and Southeast Asia generally do not have usable DHS surveys circa 2000. Thus we confine our comparison to two areas: Sub-Saharan Africa and a smaller South Asian

[16] Since high blood pressure and asthma are only available for India, we do not include them here.

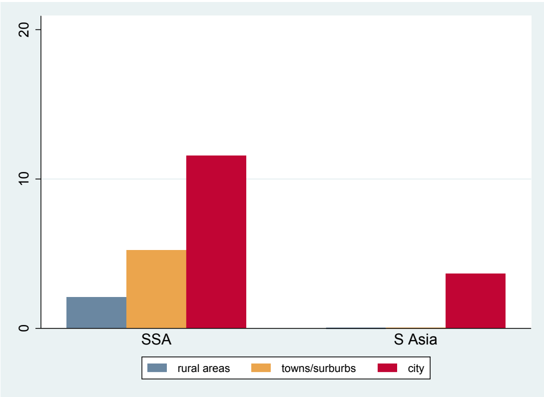
Figure 10: Percentage of households with utilities, 2000-2015



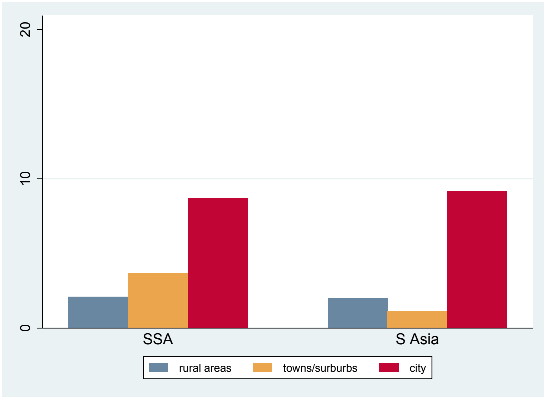
(a) Electricity 2000



(b) Electricity 2015



(c) Piped water into dwelling 2000



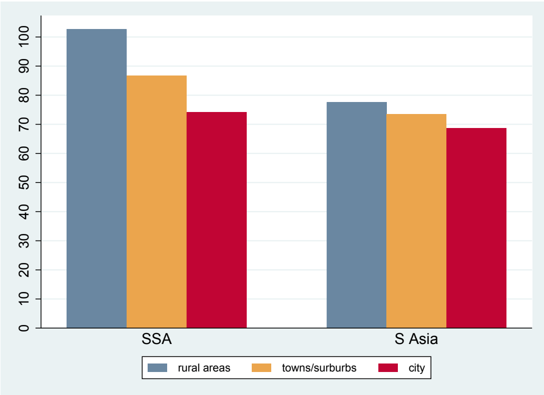
(d) Piped water into dwelling 2015

In panel (c), piped water into dwelling is unavailable for Senegal 1997, where instead respondents are asked if their main source of water is more generally ”piped into dwelling/plot/yard”.

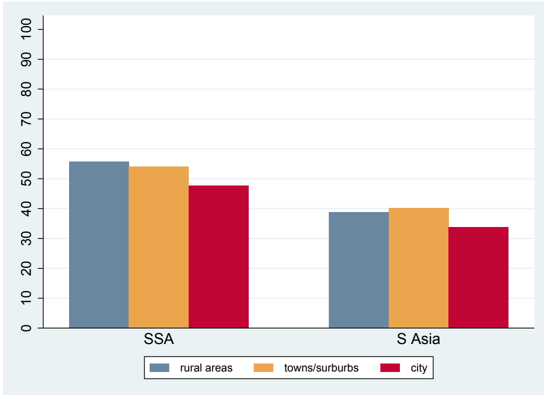
sample limited to Bangladesh and Nepal. Since the sample of countries differs from the one used in Section 5.1 for both regions, the 2015 graphs reported here will not generally match those in Section 5.1. While most outcomes get noticeably better over time, there is one exception.

We start with utilities. Electricity connections improved pretty dramatically everywhere. Rural electrification nearly doubled from low initial rates around 10% in Africa, and the urban shares also increased by about 10 percentage points from a higher base (Figure 10). Interestingly, while the South Asian pattern of worse outcomes in towns and suburbs is often true for 2015 even when excluding India, it is much less true in 2000, suggesting that rural areas saw their infrastructure and other outcomes improve even faster than towns and suburbs.

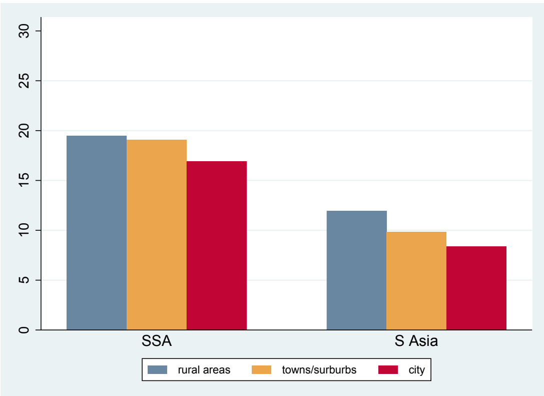
Figure 11: Health outcomes, 2000-2015



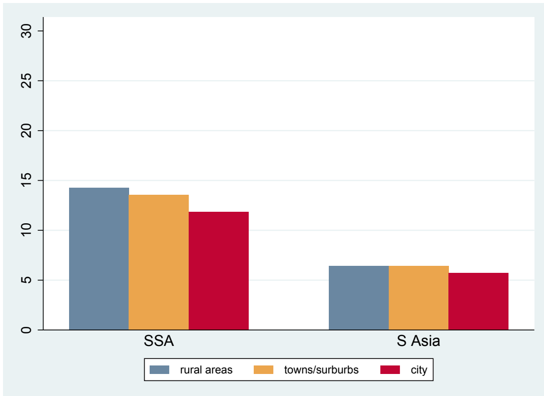
(a) Infant mortality rate 2000



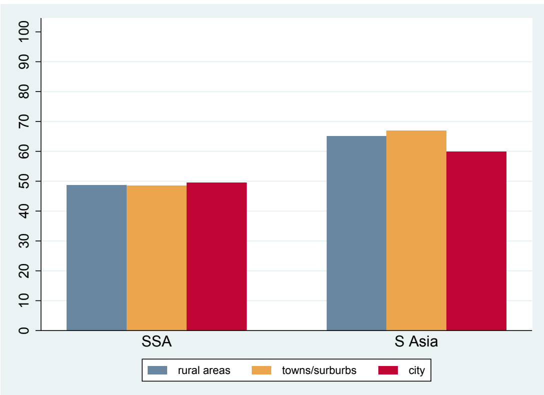
(b) Infant mortality rate 2015



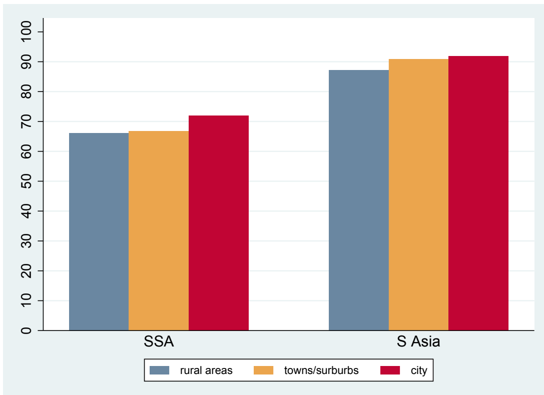
(c) Diarrheoa 2000



(d) Diarrhoea 2015



(e) Third dose of DPT vaccine 2000

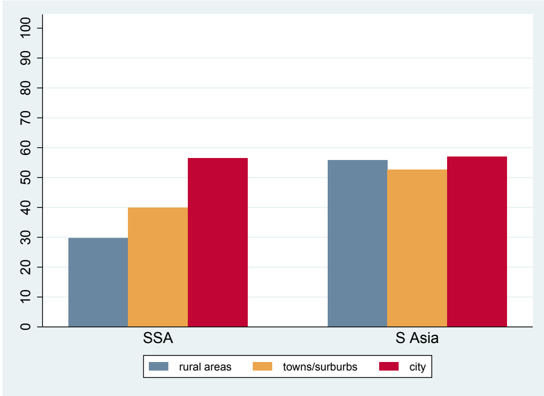


(f) Third dose of DPT vaccine 2015

In panels (a) and (b), the infant mortality rate is defined as the probability of dying before the first birthday, per 1000 live births. Panels (c) and (d) indicate the percent of of children aged 5 and under with diarrhoea in the last two weeks. Panels (e) and (f) present the percentage of children aged 12-24 months that received the third dose of the DPT vaccine.

Figure 12: Percentage of 16-year-olds with at least 8 years of education, 2000-2015

(a) 2000



(b) 2015

In panel (a), data on schooling for 16-year-olds was unavailable for Senegal 1997.

In both areas, piped water rates are under 12% everywhere in both time periods. While the piped water connections increased in South Asia (Figures 10c and 10d), piped water penetration actually fell in urban Africa, and was essentially unchanged in rural Africa. One potential explanation is that new urban residents moved to sites on the edge of cities that had been rural, with limited infrastructure, in 2000. African cities have faced a challenge of providing services to meet the needs of a growing population. Piped water, which requires laying down water mains and often retrofitting, is more expensive to provide than extending electricity lines, so the improvement in one but not the other is not surprising.

In Figure 11 we turn to the main recorded health outcomes. Infant mortality fell in our South Asian countries by about 50% in each zone. Sub-Saharan Africa saw a similar drop close to 50% in rural areas, and smaller but substantial decreases in urban areas as well (Figures 11a and 11b). Rates of diarrhoea fell less dramatically but still substantially, by about 5 percentage points across location types from base rates of 8–20 (Figures 11c and 11d). Absolute gaps across zones were thus smaller at the end of the period in both measures. DPT vaccinations (Figures 11e and 11f) saw gains of over 15 percentage points in all zones from a base of about 50-65%, with the largest gains in urban areas, which actually lagged slightly behind rural areas in Bangladesh and Nepal in 2000. All these improvements are likely a testament to broadly improved health care, as well as better education and information.

In Figures 12a and 12b, the fraction of 16-year-old children who have completed 8 years of schooling jumped everywhere by 10 to 20 percent points in Sub-Saharan Africa and by 20–30 percent points in our South Asian sample. Within regions the rural-city differential in Sub-Saharan Africa of 20 points in 2000 increased to over 25 percent points in 2015, despite some rural improvement over time. As noted above, today city rates are about double those of rural. In contrast, in South Asia the rural-city gap fell remarkably from almost 30 points, such that today rural and city rates of education are almost the same. Bangladesh and Nepal have made great strides in increasing especially rural education.

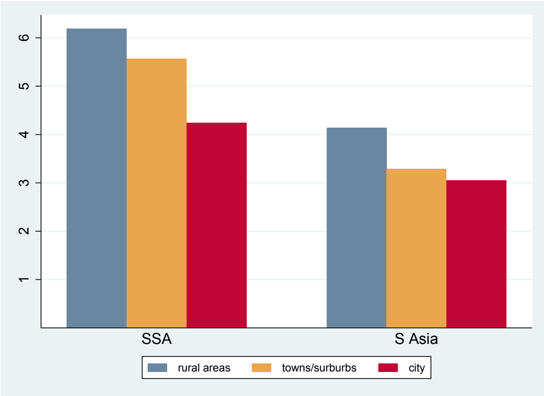
Changes in fertility and fertility preferences are reported in Figure 13. Fertility rates declined in both regions in all zones. The declines in Sub-Saharan Africa are more modest: about 1/2 a child from a base of 4-6. This is a period in which many have worried about a "stalled" demographic transition in Africa (e.g. Bongaarts and Casterline, 2013). Rates have continued to fall in all zones, but very slowly, at least in this sample of countries. To the extent that urbanisation has decreased overall rates via composition effects, the urban rate of still nearly 4 children per woman is worrying. In South Asia there are deeper declines of 0.75 to 1.5 children from a smaller base. Within regions, strong urban-rural differentials persist in Sub-Saharan Africa but have narrowed in Bangladesh and Nepal.

What may be just as discouraging is the tiny declines in desired number of children for women in Sub-Saharan Africa (Figures 13c and 13d), although male declines are larger (Figures 13e and13f). Our South Asian sample shows very modest declines for both men and women, but from a much lower base. As above in the 2015 cross section, within regions, 2000 spatial differences in actual fertility were larger than differences in desired number of children, which is relatively invariant across zones.

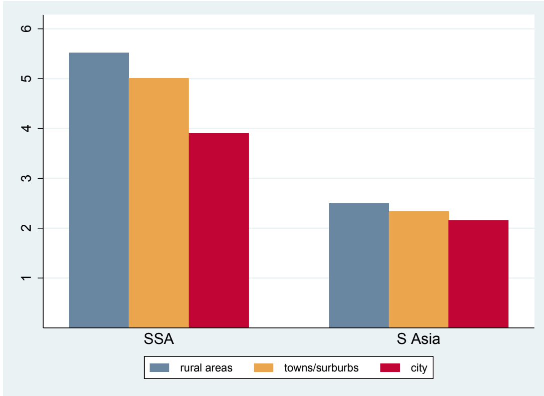
The limited changes in fertility and fertility preferences in Sub-Saharan Africa are perhaps surprising given much larger changes in the use of modern contraceptives. In Figure 14, usage among females almost doubled in rural areas and towns and rose by about 75% in cities. The increase for males is an even more striking 4-6 fold change, with the gaps in reported use between males and females declining substantially.

One potential challenge in interpreting the male changes is the nature of the question asked.

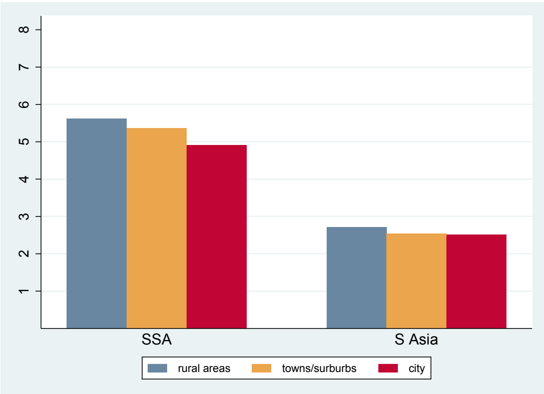
Figure 13: Fertility, 2000-2015



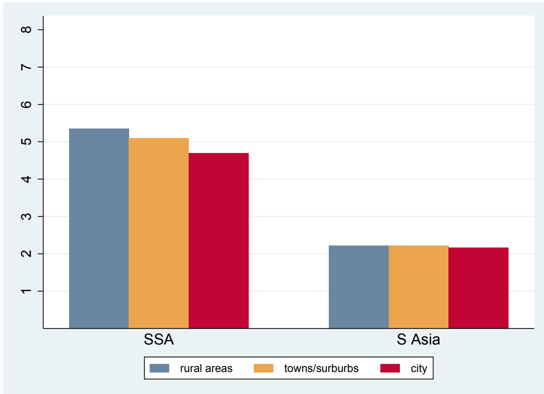
(a) Total fertility rate, 2000



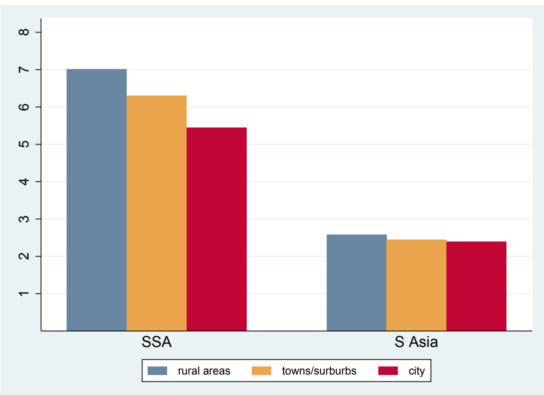
(b) Total fertility rate, 2015



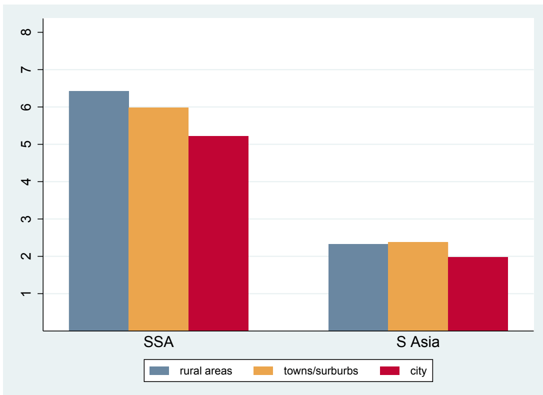
(c) Female average ideal number of children, 2000



(d) Female average ideal number of children 2015



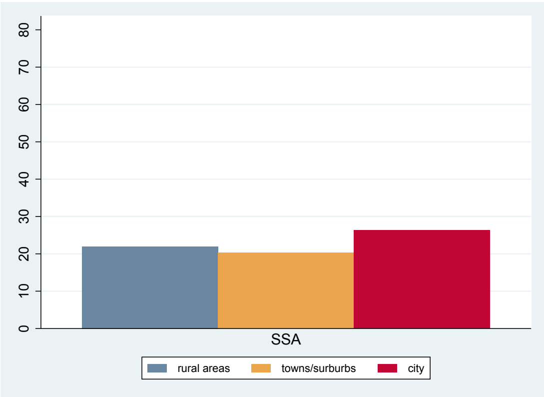
(e) Male average ideal number of children 2000



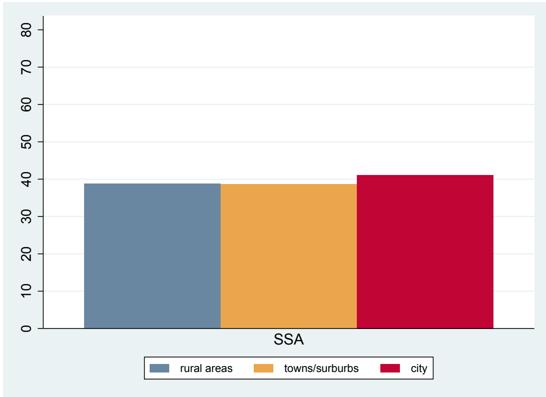
(f) Male average ideal number of children 2015

In panels (a) and (b), the total fertility rate is calculated as the number of births per thousand women if they were to give birth at the current (past 3 years) age-specific (15–19, 20–24, ...45–49) fertility rates. For panels (c)–(d), the sample is women aged 15–49. For panels (e)–(f), the sample is men aged 15–59.

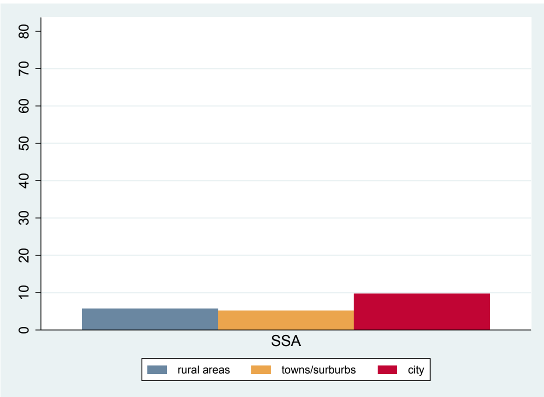
Figure 14: Percentage of respondents currently using modern contraception, 2000-2015



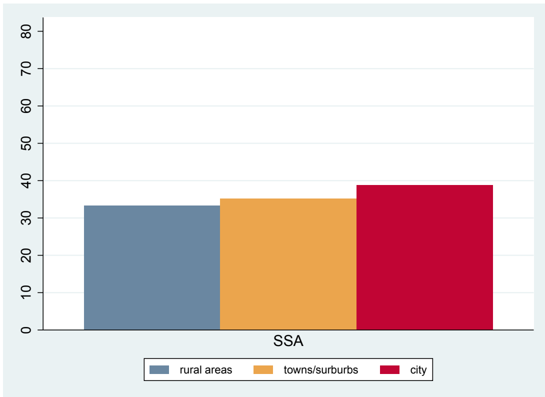
(a) Female, 2000



(b) Female, 2015



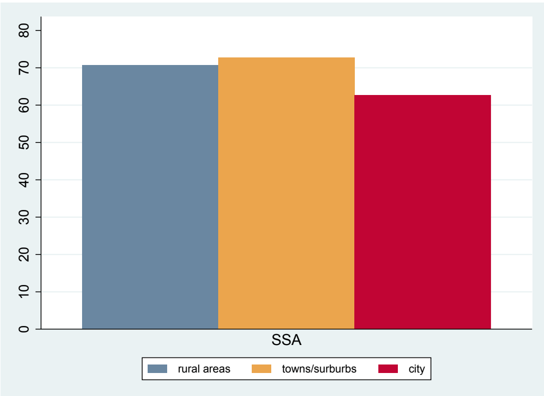
(c) Male, 2000



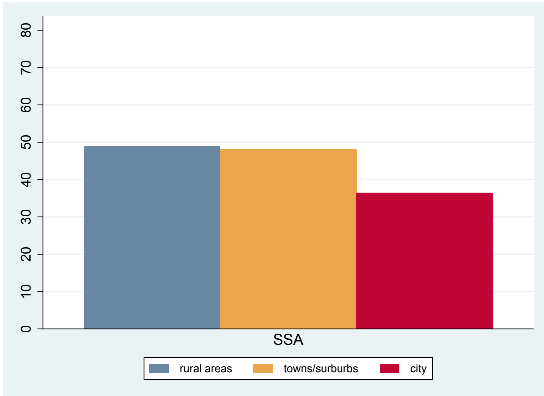
(d) Male, 2015

Modern contraception includes the following forms: pill, IUD, injections, diaphragm, condom, female sterilisation, male sterilisation, implants, female condom, foam and jelly, emergency contraception, other modern. In panels (a) and (b), data on female modern contraception use was missing for: Bangladesh 1999–00, Tanzania 1999, Burkina Faso 1998–99, Côte d'Ivoire 1998–99, Ghana 1998, Guinea 1999, Senegal 1997, Togo 1997, and Guinea 2012. In panels (c) and (d), data on male modern contraception use was missing for: Kenya 2003, Tanzania 1999, Uganda 2000-01, Namibia 2001, Burkina Faso 1998–99, Côte d'Ivoire 1998–99, Ghana 1998, Guinea 1999, Senegal 1997, Togo 1997, and Guinea 2012.

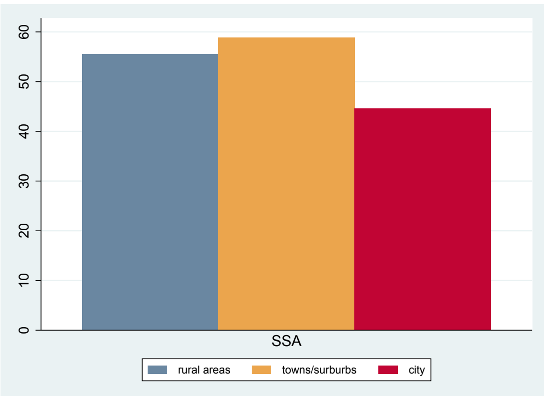
Figure 15: Percentage of respondents who believed wife beating was justified for at least one reason, 2000–2015



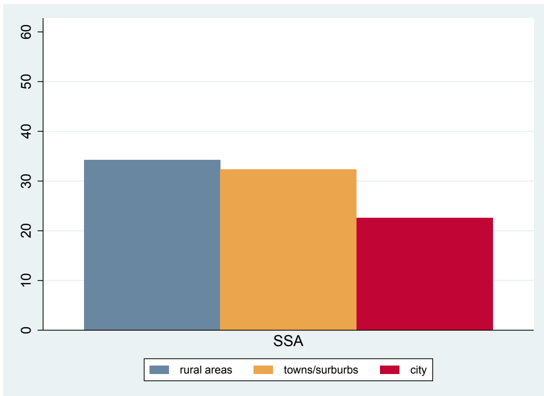
(a) Female, 2000



(b) Female, 2015



(c) Male, 2000



(d) Male, 2015

Panels (a)–(d) plot the percentage of respondents (aged 15–49) whose opinion is that a husband is justified in hitting or beating his wife when: she goes out without telling him, she argues with him, she refuses to have sex with him, or she burns the food. Data was missing for the following surveys for panels (c) and (d): Tanzania 1999, Burkina Faso 1998–99, Côte d'Ivoire 1998–99, Ghana 1998, Guinea 1999, Senegal 1997, Togo 1998. Panel (a) is missing for Namibia 2000, and panel (c) is further missing for Zimbabwe 1999.

Men are asked whether they use contraception as a couple, regardless of who is using it. Thus, it is possible that part of the improvement for males is a better understanding of family planning and their roles in joint decisions to practice birth control.

Finally, in Figure 15, we turn to violence in Sub-Saharan Africa. We can only compare attitudes, not victimisation, over time. Rates of acceptance of wife beating fell roughly in parallel across location categories for both men and women, but substantially so, by over 20 percentage points from bases of about 50 (male) and 70 (female).

5.4 Accounting for sorting

The differentials we have seen could be due in large part to the fact that people who live in cities are different on average from those who live in rural areas on observable dimensions like education and age, and unobservable ones like motivation and various abilities. These are sometimes known as composition effects, and when they affect choices such as where to live, they are known as sorting. This is a key concern in economic studies (see Altonji and Mansfield (2018) for a review of the literature).

As noted above, the implications for health outcomes are clear: if more educated people know more about how to protect themselves and their children from illness, controlling for education allows us to understand whether city-rural differentials are due only to the presence of a more-educated population, or to something else. But this logic extends to all outcomes we study, including crime, where education functions more as a rough proxy for earning potential that is largely fixed for adults. For example, highly educated people in cities and rural areas are both more likely than their less educated counterparts to have property to protect and the means to protect it, including sturdier homes, better locks, and domestic employees like security guards. They may also be better able to protect themselves, by living in well-lit neighbourhoods and travelling in private cars or taxis, rather than walking at night. Within cities they may not bear the brunt of crime.

This suggests that cities could have lower crime rates than rural areas because they have more people who have the means to protect themselves from it, not because the city is safer per se. Controlling for household and individual characteristics allows us to make comparisons between urban and rural people who are similar along observable dimensions.

Correcting for sorting in this way is far from perfect. There are many unobserved attributes such as wealth and unobservable attributes such as motivation, for which we cannot control. The general presumption is that we thus under-correct for sorting.

We present results in this section with controls on individual and household characteristics to correct for sorting as best we can. Appendix Tables C.1–C.10 report regression results, based on equation (3) for fifteen key outcomes circa 2015. Results are restricted to Sub-Saharan Africa and South Asia. In each table, for each region, we first report estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, the differentials of cities and town/suburbs relative to the rural baseline, with no controls except for a full set of country fixed effects. In other words, we do not include the X_{ijc} term in that first column. The results are generally similar to the differentials between rural and the other two categories shown in the corresponding figures in Section 5.1, with differences for two reasons. First, the country fixed effects sweep out average differences in outcomes across countries. Second, we use different weights. In the graphs in Section 5.1 we weighted average outcomes for each zone type (city, town, rural) by each country's share of that area type's regional population, in addition to the DHS sampling weights. Here we use weights more appropriate for these regressions, defined in equation (4). Besides the DHS weight, we weight by the share of each country's population relative to its survey sample size, to diminish the role of countries with larger surveys relative to their population.

In the Appendix tables, in the second columns, we then report estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ using the full specification with controls for sex, age, and education of the household head, and of the person in question and their mother where applicable. These controls remove the effect of sorting by these observable attributes, accounting for the fact that different types of people live in cities, towns, and rural areas as noted above. If city children are only more likely to be vaccinated because their mothers are more educated, but children with mothers of the same education status are equally likely to be vaccinated in urban and rural areas, this second regression result will tell us that. Of course, as noted above there could be sorting on wealth, motivation, genetic disposition to disease, or many other factors we cannot measure. We do see that accounting for age, sex and education alone removes a substantial share of the urban-rural differential for many outcomes, but typically not all of it. Education of both the mother and household head play a strong role in most outcomes.

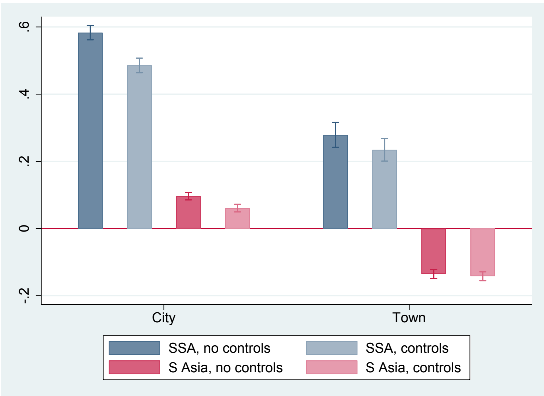
Estimates of the city-rural and town-rural differentials ($\hat{\beta}_0$ and $\hat{\beta}_1$) are also presented here in graph form, with 95% confidence intervals. Each graph reports city-rural differentials on the left and town-rural on the right. In each part we first present Sub-Saharan Africa with no controls and then add controls for sorting; in the next two bars we do the same for South Asia. Thus the left side of each graph has four bars for cities and the right side four for towns. In the discussion to follow we focus on the city-rural differentials. In Sub-Saharan Africa, the town-rural differentials tend to have the same sign as city-rural, but smaller magnitudes. As we have seen already, South Asian towns tend to have worse outcomes than their rural counterparts in most cases.

In the discussion we focus on how the differentials between city and rural are changed by sorting. So if we have a coefficient of -0.0095 (approximately the case for Sub-Saharan Africa) for reduced likelihood of infant mortality for city without sorting controls and -0.00045 with controls, we can say that the city-rural differential falls by about 1/2. However it is also important to think about this differential relative to the mean infant mortality rate. Say the mean infant mortality rate is 0.05, which is approximately correct for Sub-Saharan Africa. Then cities have about a 20% (0.0095./0.05) lower infant mortality rate which falls after accounting for sorting to under a 10% differential– still arguably non-trivial. If, however, the base infant mortality rate were instead a horrible 0.40 then prior to accounting for sorting cities would have have only a 2.5% differential (.0095/0.40), which after sorting would shrink to a 1.25% differential. These are much more modest numbers in the face of the horrendous 40% infant mortality rate. While we focus on the relative drop in differential (in the example roughly 50%), we also comment on the overall rates and the degree to which changing differentials might be viewed as important. These relevant mean (for the sample with sorting) rates are reported in the notes to all figures.

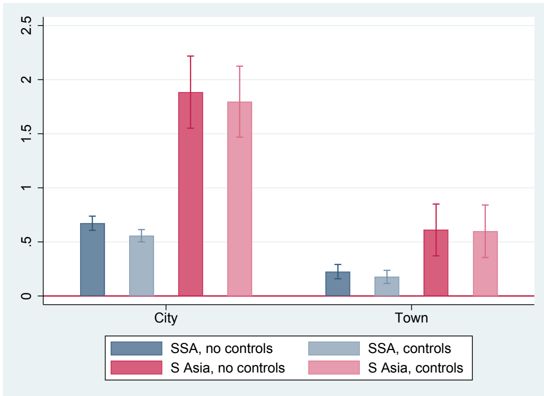
Sorting accounts for a relatively small share, typically less than 20%, of the large city-rural and smaller town-rural differentials in electricity (Figure 16a) and safe drinking water (Figure 16b).

We note the absolute city-rural differentials for Sub-Saharan Africa are very large relative to the mean level of servicing, indicating the importance of space on servicing. The modest shrinking of differentials with sorting accounted for suggests that differences in utilities are more supply-driven, and less based on differences in household demand and composition of families across space. While cost must be a factor in the underlying supply decisions to have less supply in more sparsely populated areas, this lack of supply could in turn be a factor inducing migration, further depopulating these areas.

Figure 16: Percentage of households with utilities



(a) Electricity



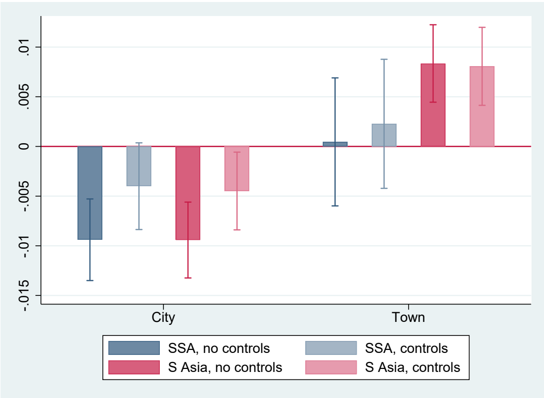
(b) Safely managed drinking water

The outcome means are as follows: in panel (a) for SSA 0.357 and for S Asia 0.849; in panel (b) for SSA 0.204 and for S Asia 0.647. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.1 and C.2 for all numerical values and lists of controls. Safely managed drinking water is defined by the DHS-WHO Joint Monitoring Program as all improved water sources that take zero minutes to collect, or are on the premises. Improved water sources encompass all piped water and packaged water, as well as protected wells or springs, boreholes, and rainwater.

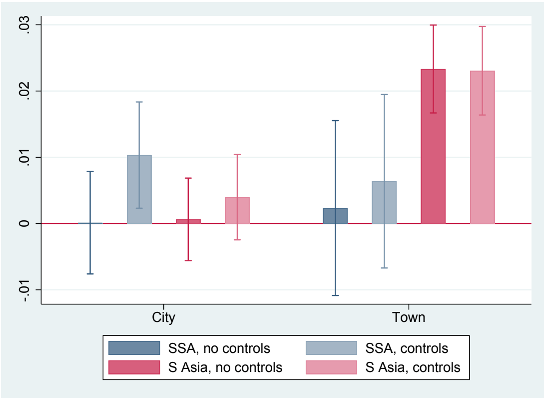
Figure 17 shows basic health outcomes. In terms of lower infant mortality in cities over rural areas, controls (such as sex of the child) reduce city-rural differentials by half in both regions, leaving only a significant effect for South Asia. In South Asia, the town penalty of higher infant mortality than rural is unaffected by sorting.

Our chief morbidity measure, children's diarrhoea in the past two weeks (Figure 17b), shows no raw city-rural differentials in either region. However, controlling for sorting in Sub-Saharan Africa implies that urban children have a 0.01 greater chance of diarrhoea than their rural counterparts. This is surprising given the more widespread access to safe water and improved sanitation. It suggests that these factors are outweighed by the denser disease environment. Note that the problematic higher diarrhoea rates in South Asian towns relative to rural are

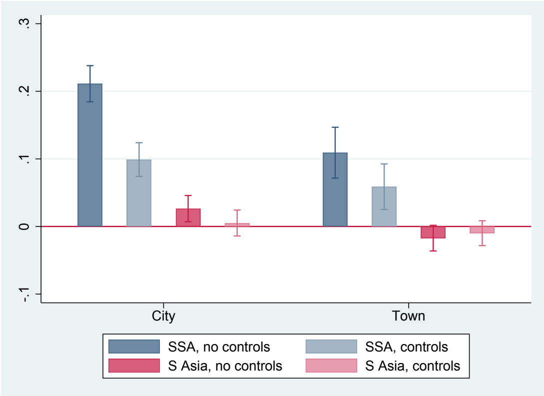
Figure 17: Main health outcomes



(a) Infant death



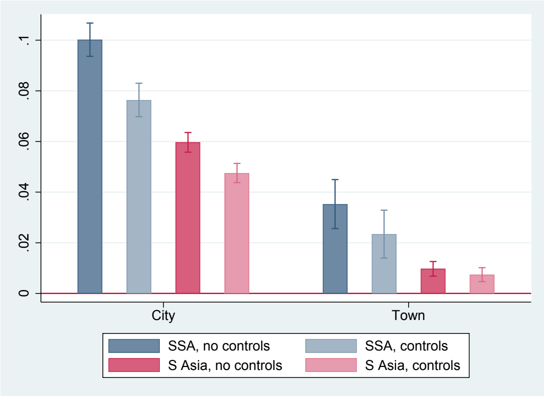
(b) Diarrhoea



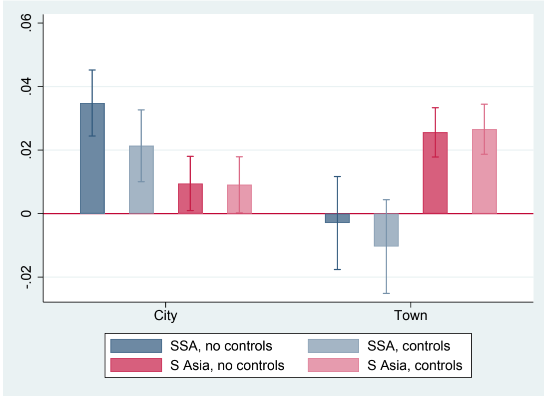
(c) Third dose of DPT vaccine

The outcome means are as follows: in panel (a) for SSA 0.053 and for S Asia 0.04; in panel (b) for SSA 0.142 and for S Asia 0.089; in panel (c) for SSA 0.642 and for S Asia 0.802. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.3–C.5 for all numerical values and lists of controls. In panel (a), the dependent variable infant death is a dummy for whether or not the child died within 1 year of being born, for children who were born between 1 and 5 years before the survey. Panel (b) indicates the percent of children aged 5 and under with diarrhoea in the last two weeks. Panel (c) presents the percentage of children aged 12-24 months that received the third dose of the DPT vaccine.

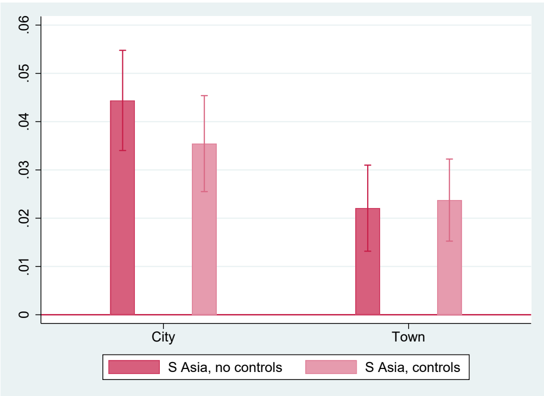
Figure 18: Other health outcomes



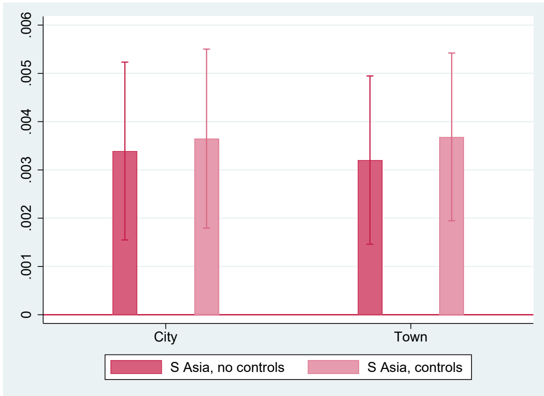
(a) Obesity



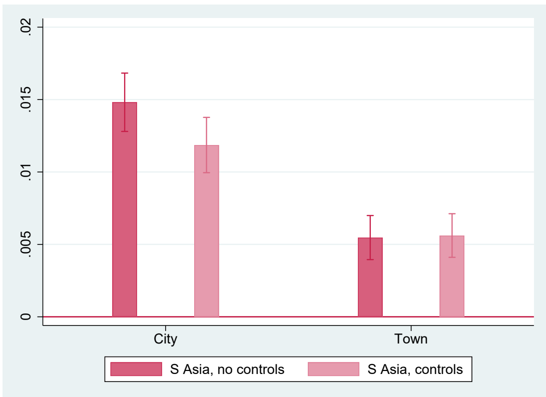
(b) Cough



(c) High blood pressure, India 2015–16



(d) Asthma, India 2015–16



(e) Diabetes, India 2015–16

The outcome means are as follows: in panel (a) for SSA 0.088 and for S Asia 0.053; in panel (b) for SSA 0.21 and for S Asia 0.137; in panel (c) 0.245; in panel (d) 0.018; in panel (e) 0.016. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.9–C.8 for all numerical values and lists of controls. In panel (a), obesity is defined as having a body mass index ≥ 30 for 20–49-year-olds. The sample consists of household members that were 20–49 years old, eligible for the individual male and female surveys, and not pregnant. Data on obesity for the Angola 2015–16 survey was unavailable. Panel (b) displays results for children aged 5 and under who have had a cough in the last two weeks. Panel (c) shows results for household members aged 25 and over with high blood pressure. High blood pressure is defined as a mean systolic measure ≥ 140 or a mean diastolic measure ≥ 30 over three blood pressure readings. Respondents who reported taking medication for hypertension, or who reported that a medical professional told them they were hypertensive, were also coded as having high blood pressure. Panels (d) and (e) display results for 15–49-year-olds that self-reported asthma and diabetes, and who were eligible for the individual male and female surveys. In panel (d) for diabetes, pregnant women are additionally excluded from the sample.

unaffected by sorting.

Finally, controlling for sorting reduces the large positive city-rural differential (and the town-rural differential) in DPT-3 vaccinations in Sub-Saharan Africa by over 50%, and eliminates the small urban advantage in South Asia (Figure 17c). This suggests that cities don't just see higher rates just because they offer better and more health facilities and easier access to vaccines; they also do so because they have more educated mothers and household heads. In general in these regions, we note that even the third shot of DPT vaccination is widely sought everywhere, with only children in the least educated families having noticeably lower rates.

In Figure 18, we examine health problems related directly to urban pollution, diet, and stress, all of which show higher rates in cities. The coefficients, given base rates, imply that adult obesity is more than twice as high in cities as in rural areas in both Sub-Saharan Africa and South Asia. Sorting explains a only about 1/4 of this differential. The corresponding differential in children's cough is in the same direction but much smaller. In India, urban residents are twice as likely as their rural counterparts to have diabetes given the base rates, and 15–20% more likely to have asthma or high blood pressure. Controlling for sorting modestly increases the city-rural differentials for asthma, but for all other health outcomes reduces them, with about a 20% reduction in the differential. The regression results in Appendix Tables C.6–C.10 show that having completed primary school raises the incidence of these health conditions relative to having not completed primary school. In most cases the secondary and college educated also have higher rates than the primary educated. In the case of children's cough, these differences might be influenced by possibly greater reporting propensity of higher educated families, rather than actual incidence. However, reporting propensities cannot explain the measured differentials for obesity and high blood pressure which are measured by interviewers. All this suggests that not only are these urban ills, but they are also ills of the more well-to-do, potentially because of a diet with more salt, fat, sugar, and processed grains. However, cities also have fewer underweight (BMI ≤ 18.5) adults as well, so the diet picture is more nuanced than we explore here.[17]

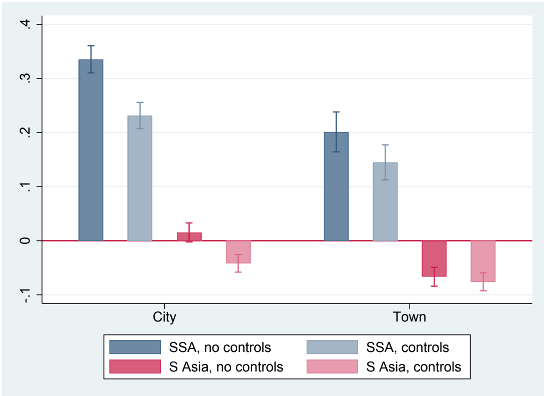
[17] For example, the city-rural differential on being underweight in Sub Saharan Africa is 0.053 which shrinks to 0.033 after sorting is accounted for, with an overall mean of 0.103.

Figure 19a shows that the huge city-rural differences in educational attainment in Sub-Saharan Africa are not all due to availability of schooling and probably better staffing. Sorting accounts for roughly a third of the city-rural and town-rural differentials in the educational attainment of 16-year-olds. In South Asia, these controls actually reverse the city-rural differential: controlling for education and age of the household head, rural 16-year-olds are more likely to have attained 8 years of schooling than their city counterparts, although the differential is small (4 percentage points with a high mean). This may suggest that access to education in cities is a problem for some families, or that the opportunity cost of having children in school in South Asian cities is high.

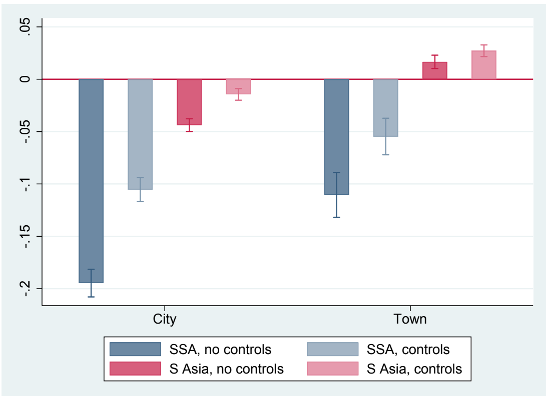
Turning to fertility in Figure 19b, sorting accounts for about half of the city- and town-rural differentials in Sub-Saharan Africa. However, even accounting for this sorting, city rates are about 25% lower than rural, given the base rates. There remains some prospect that urbanisation will decrease fertility at least in Sub-Saharan Africa. There are several possible channels as noted earlier, including greater availability of contraception, a higher opportunity cost of children, and social pressures affecting the acceptance of family planning. In South Asia, sorting accounts for two-thirds of the raw city-rural differential, and the remaining difference represents less than 10% of the mean. Patterns in the use of modern contraception in Sub-Saharan Africa are similar to fertility patterns. Sorting accounts for about 55% and 45% of the city-rural and town-rural differentials (Figure 19c). In South Asia, in contrast to other regions as noted above, use of modern contraception is lower in cities and towns compared to the countryside; controlling for sorting slightly increases that differential.

Finally, we turn to crime and violence. First, in Figure 20, we examine attitudes toward and experience of domestic violence in Sub-Saharan Africa and South Asia. In Sub-Saharan Africa (Figure 20b), 35–50% of the city-rural and town-rural differentials in women who think wife beating is justified for any reason can be explained by sorting, but in Figure 20a sorting cannot explain lower city and town rates of experienced violence. In contrast in South Asia, lower rates in cities compared to the countryside for both attitudes and experience of violence are both driven entirely by sorting. As usual in South Asia, towns look worse than rural areas, and this is not explained by sorting.

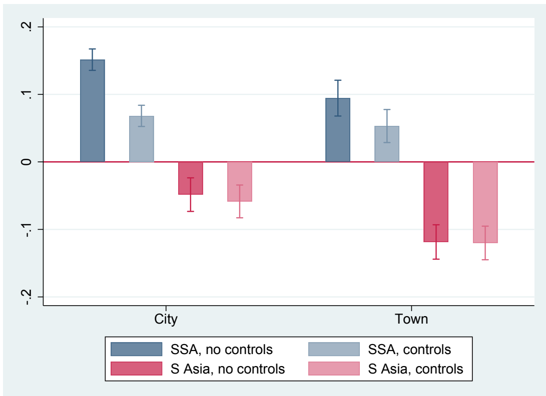
Figure 19: Education and Fertility



(a) 8 years of schooling among 16-year-olds



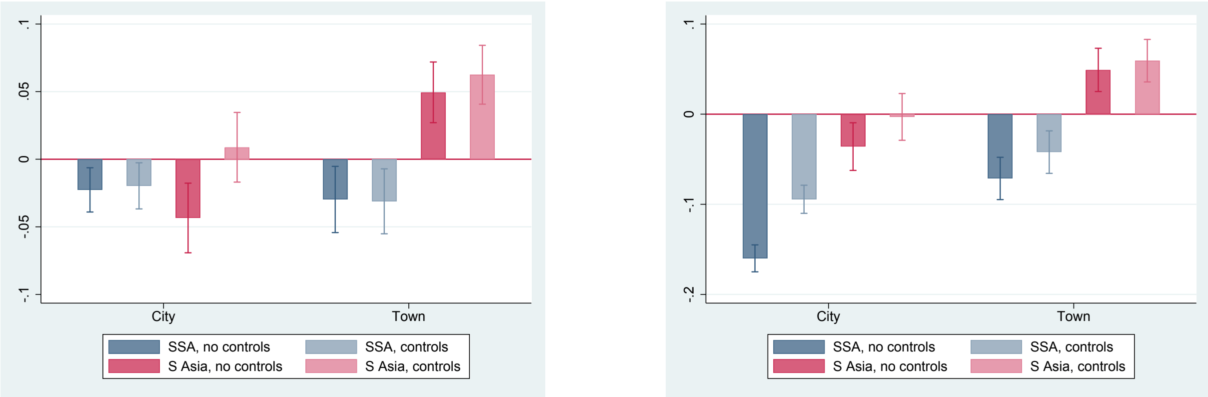
(b) Number of births in past three years



(c) Current use of modern contraception, female

The outcome means are as follows: in panel (a) for SSA 0.363 and for S Asia 0.765; in panel (b) for SSA 0.481 and for S Asia 0.225; in panel (c) for SSA 0.331 and for S Asia 0.566. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.11–C.13 for all numerical values and lists of controls. Panel (b) presents results the total number of births in the past three years for women aged 15–49. Panel (c) presents results for women aged 20–40 using modern contraception. Modern contraception includes the following forms: pill, IUD, injections, diaphragm, condom, female sterilisation, male sterilisation, implants, female condom, foam and jelly, emergency contraception, other modern.

Figure 20: Domestic Violence



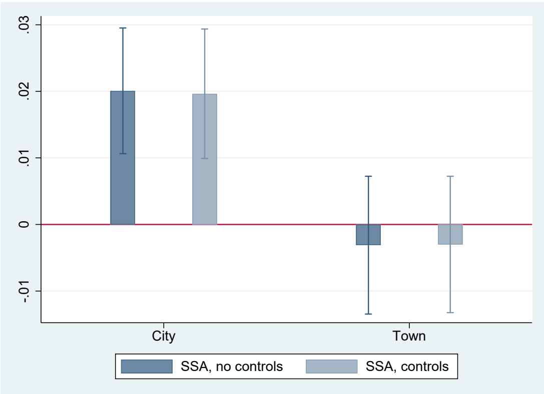
(a) Currently married women who have ever experienced physical spousal violence

(b) Women who think beatings are justified

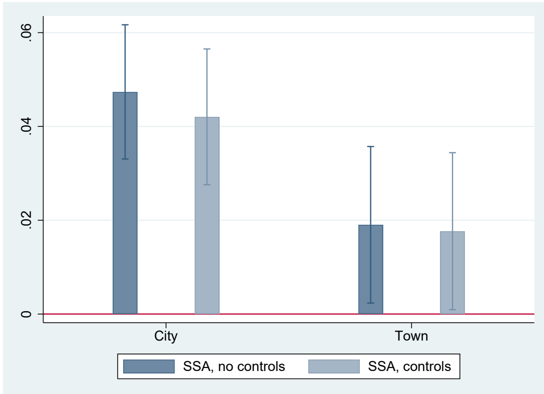
The outcome means are as follows: in panel (a) for SSA 0.296 and for S Asia 0.306 and in panel (b) for SSA 0.485 and for S Asia 0.429. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.14–C.15 for all numerical values and lists of controls.

As noted in Section 5.1, incidence and fear of physical and property crime, as reported by Afrobarometer, are higher in cities than in rural areas. Figure 21 shows that sorting plays essentially no role in explaining either the city-rural differentials or the smaller and noisier town-rural differentials. Nevertheless, like with actual incidence of domestic violence, after sorting city-rural differentials are fairly small: urban households are less than 10% more likely to experience physical attacks and robberies than their rural counterparts, given the base rates.

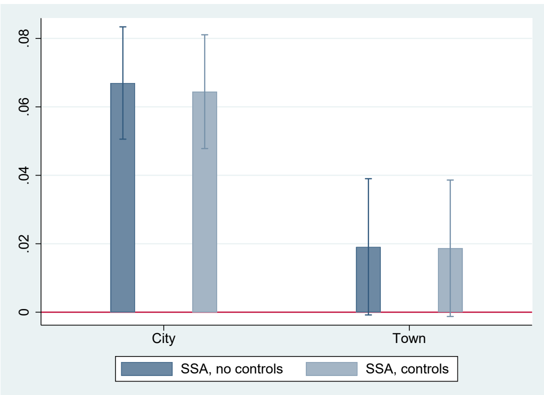
Figure 21: Crime



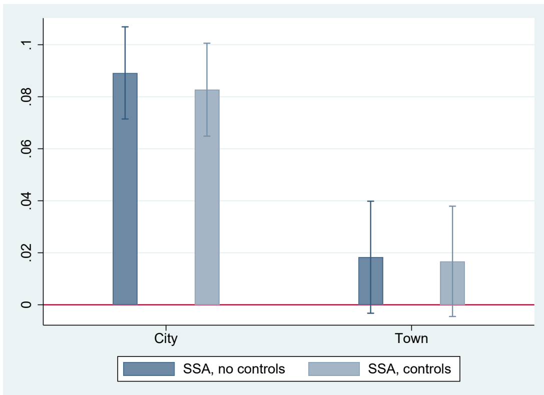
(a) Households with family members that have been physically attacked in the past year



(b) Households that had something stolen from their house in the past year



(c) Households with family members that have feared crime in their own home in the past year



(d) Households with family members that have felt unsafe walking in their neighbourhood in the past year

The outcome means are as follows: in panel (a) 0.094, in panel (b) 0.278, in panel (c) 0.316, and in panel (d) 0.373. The bars are estimates of β_0 and β_1 from equation (3) and the whiskers indicate the 95% confidence intervals. The omitted category is rural. See Tables C.14–C.15 for all numerical values and lists of controls.

5.5 Differences by political and administrative status

Tables 1–5 report regression results based on a variant of equation (3), adding controls for clusters' political status, administrative urban-rural status, and local density, these replacing controls in the prior section for individual- and household-level characteristics. Using point data on national and regional capitals from Statoids,[18] we assign DHS clusters as having no political status, regional capital status, or national capital status.[19]

For each set of outcomes, in the first column we have town and city as indicator variables. In the second column, we add designations if the cluster is in a national capital or in a regional capital. In the third column we add an indicator of the cluster being in a DHS urban defined area (usually based on the country's census designation). This variable may represent urban administrative status, or may capture urban character in another sense, depending on the national definition. For example in Africa we might think urban status in Francophone counties is an administrative designation, while in Anglophone countries it may be much broader adding in places that are not cities per se but are urban in character. Finally in the fourth column, we include average population density within a fixed radius of each DHS cluster (2km for urban clusters and 5km for rural).

Overall, political and administrative status explains a substantial portion of city-rural and town-rural differentials, but they generally remain significant and in the same direction as the raw differential (with cities showing the advantage). The very strong effect for being located in a national capital for many outcomes may reflect a political bias in allocation of resources. Once we further control for local population density, the town-rural differential disappears for many outcomes, suggesting that we cannot distinguish town status from higher population density in a more continuous sense. Controlling for local population density reduces substantially, but does not eliminate, city-rural differential for many outcomes. The fact that density can explain much of the non-administrative and non-political city-rural differential is not surprising since cities are defined by their high density.

[18] http://www.statoids.com/

[19] See the section 3.4 for the methodology used to assign DHS clusters political status.

In results not shown, we find that sorting again generally tends to mute effects but not eliminate them. We additionally tried many interaction effects; for example, being in national capital buffer and city polygon versus being in a national capital buffer and town, or national capital buffer and rural polygon. Interactions convey more complex patterns, but generally do not dominate or change conclusions.

More specifically, in Table 1, after controlling for political, administrative, and urban status as well as population density, access to utilities still remains roughly 10%–14% higher in cities than in rural areas. The results also indicate that living in a national capital gives hugely improved access to improved sanitation, electricity, and drinking water. Further, urban status seems to matter more than regional capital status. This is perhaps not surprising since some regional capitals are towns, not cities, in our typology. Adding local population density in columns 4, 8, and 12 has particularly large effects relative to the base (cols 1, 5, and 9): town effects are weakened or disappear entirely.

In Tables 2 and 3, health outcomes follow particular patterns. In Table 2, obesity is worse in cities, and even more so in national capitals, as well as in clusters designated as DHS urban and in regional capitals. Vaccination rates remain higher in cities after adding all controls. For diarrhoea and cough in Table 3 as well as the third DPT vaccination in Table 2, we find that being in a national capital is actually worse for health outcomes. One speculation is that capitals, as large hubs of migrants, have a higher presence of slums, as well as pollution.

In Table 4, again being in a national capital, being DHS urban, and having high density are important for schooling and fertility. For modern contraception use, population density (column 12) drowns out all other relationships except urban administrative status.

Finally in Table 5 columns 2–4, we see again that being in a national capital, local density, and DHS urban status play a large role for attitudes towards domestic violence. In columns 6-8, results on actual physical domestic violence are less conclusive, perhaps because of lower rates overall.

Table 1: Utilities

	(1) Improved sanitation	(2) Improved sanitation	(3) Improved sanitation	(4) Improved sanitation	(5) Electricity	(6) Electricity	(7) Electricity	(8) Electricity	(9) Safe water	(10) Safe water	(11) Safe water	(12) Safe water
City	0.478***	0.377***	0.223***	0.103***	0.583***	0.495***	0.280***	0.140***	0.359***	0.291***	0.160***	0.090***
	(0.010)	(0.013)	(0.014)	(0.018)	(0.011)	(0.015)	(0.016)	(0.022)	(0.009)	(0.013)	(0.012)	(0.016)
Town	0.175***	0.165***	0.096***	0.010	0.279***	0.271***	0.174***	0.078***	0.133***	0.127***	0.068***	0.015
	(0.016)	(0.016)	(0.016)	(0.017)	(0.019)	(0.019)	(0.017)	(0.021)	(0.011)	(0.011)	(0.010)	(0.013)
National Capital		0.235***	0.170***	0.157***		0.253***	0.163***	0.150***		0.199***	0.144***	0.134***
		(0.015)	(0.015)	(0.015)		(0.016)	(0.014)	(0.014)		(0.017)	(0.016)	(0.017)
Regional Capital		0.148***	0.088***	0.078***		0.097***	0.013	0.003		0.071***	0.020	0.015
		(0.012)	(0.012)	(0.012)		(0.014)	(0.013)	(0.013)		(0.012)	(0.012)	(0.013)
DHS Urban			0.244***	0.206***			0.339***	0.310***			0.207***	0.198***
			(0.013)	(0.015)			(0.014)	(0.016)			(0.010)	(0.012)
ln(pop density)				0.029***				0.032***				0.015***
				(0.004)				(0.004)				(0.003)
Region	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA
Outcome mean	0.451	0.451	0.451	0.471	0.357	0.357	0.357	0.382	0.204	0.204	0.204	0.221
R^2	0.272	0.286	0.312	0.309	0.391	0.405	0.460	0.455	0.207	0.218	0.247	0.247
N	413313	413313	413313	374983	413359	413359	413359	374998	413575	413575	413575	375191
Surveys	29	29	29	29	29	29	29	29	29	29	29	29

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table 2: Health outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Obese	Obese	Obese	Obese	DPT 3	DPT 3	DPT 3	DPT 3
City	0.100***	0.076***	0.044***	0.032***	0.211***	0.209***	0.140***	0.083**
	(0.003)	(0.004)	(0.005)	(0.006)	(0.014)	(0.018)	(0.021)	(0.026)
Town	0.035***	0.033***	0.019***	0.011	0.109***	0.108***	0.079***	0.034
	(0.005)	(0.005)	(0.005)	(0.006)	(0.019)	(0.019)	(0.019)	(0.022)
National Capital		0.058***	0.044***	0.041***		-0.048**	-0.080***	-0.076***
		(0.006)	(0.006)	(0.006)		(0.017)	(0.017)	(0.018)
Regional Capital		0.032***	0.019***	0.017***		0.038*	0.009	0.005
		(0.005)	(0.005)	(0.005)		(0.017)	(0.016)	(0.016)
DHS Urban			0.051***	0.046***			0.114***	0.096***
			(0.004)	(0.005)			(0.014)	(0.018)
ln(pop density)				0.003**				0.011*
				(0.001)				(0.004)
Region	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA
Outcome mean	0.088	0.088	0.088	0.091	0.642	0.642	0.642	0.657
R^2	0.049	0.051	0.055	0.053	0.211	0.213	0.218	0.213
N	227706	227706	227706	205873	55768	55768	55768	49893
Surveys	28	28	28	28	29	29	29	29

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level.

P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table 3: Health outcomes

	(1) Diarrhoea	(2) Diarrhoea	(3) Diarrhoea	(4) Diarrhoea	(5) Cough	(6) Cough	(7) Cough	(8) Cough
City	0.000	-0.004	-0.000	0.001	0.035***	0.016*	0.009	0.003
	(0.004)	(0.005)	(0.005)	(0.008)	(0.005)	(0.007)	(0.008)	(0.011)
Town	0.002	0.002	0.004	0.006	-0.003	-0.004	-0.007	-0.012
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.009)
National Capital		0.014*	0.015*	0.015*		0.054***	0.051***	0.049***
		(0.007)	(0.007)	(0.007)		(0.009)	(0.009)	(0.009)
Regional Capital		0.004	0.006	0.006		0.024***	0.021**	0.021**
		(0.005)	(0.005)	(0.005)		(0.007)	(0.007)	(0.007)
DHS Urban			-0.006	-0.006			0.011*	0.010
			(0.004)	(0.005)			(0.006)	(0.007)
ln(pop density)				-0.000				0.002
				(0.001)				(0.002)
Region	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA
Outcome mean	0.142	0.142	0.142	0.142	0.208	0.208	0.208	0.209
R^2	0.014	0.014	0.014	0.014	0.070	0.070	0.070	0.070
N	297759	297759	297759	265176	297866	297866	297866	265284
Surveys	29	29	29	29	29	29	29	29

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level.

P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table 4: Fertility and education outcomes

	(1) 8 years of schooling	(2) 8 years of schooling	(3) 8 years of schooling	(4) 8 years of schooling	(5) Number of births	(6) Number of births	(7) Number of births	(8) Number of births	(9) Modern contra-ception	(10) Modern contra-ception	(11) Modern contra-ception	(12) Modern contra-ception
City	0.335***	0.287***	0.160***	0.061**	-0.195***	-0.168***	-0.105***	-0.054***	0.151***	0.135***	0.073***	0.023
	(0.013)	(0.017)	(0.017)	(0.023)	(0.007)	(0.009)	(0.010)	(0.013)	(0.008)	(0.011)	(0.013)	(0.019)
Town	0.201***	0.198***	0.140***	0.070**	-0.110***	-0.108***	-0.080***	-0.044***	0.094***	0.093***	0.065***	0.024
	(0.019)	(0.019)	(0.018)	(0.022)	(0.011)	(0.011)	(0.011)	(0.012)	(0.014)	(0.014)	(0.013)	(0.017)
National Capital		0.129***	0.075***	0.059**		-0.078***	-0.052***	-0.048***		0.030*	0.007	0.008
		(0.020)	(0.020)	(0.020)		(0.009)	(0.009)	(0.008)		(0.015)	(0.015)	(0.016)
Regional Capital		0.065***	0.016	0.007		-0.024**	-0.000	-0.000		0.031**	0.008	0.008
		(0.016)	(0.016)	(0.016)		(0.008)	(0.008)	(0.008)		(0.011)	(0.011)	(0.011)
DHS Urban			0.203***	0.178***			-0.098***	-0.088***			0.096***	0.085***
			(0.014)	(0.016)			(0.007)	(0.009)			(0.009)	(0.011)
ln(pop density)				0.023***				-0.010***				0.009**
				(0.004)				(0.002)				(0.003)
Region	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA
Outcome mean	0.363	0.363	0.363	0.380	0.481	0.481	0.481	0.470	0.331	0.331	0.331	0.344
R^2	0.177	0.181	0.201	0.192	0.029	0.030	0.033	0.035	0.182	0.182	0.187	0.185
N	39133	39133	39133	35604	417251	417251	417251	381190	123944	123944	123944	112431
Surveys	29	29	29	29	29	29	29	29	28	28	28	28

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table 5: Victimization

	(1) Justified beating	(2) Justified beating	(3) Justified beating	(4) Justified beating	(5) Experienced physical violence	(6) Experienced physical violence	(7) Experienced physical violence	(8) Experienced physical violence
City	-0.160***	-0.127***	-0.073***	-0.018	-0.023**	-0.024*	-0.015	-0.004
	(0.008)	(0.010)	(0.012)	(0.015)	(0.008)	(0.009)	(0.011)	(0.016)
Town	-0.071***	-0.069***	-0.045***	-0.009	-0.030*	-0.030*	-0.027*	-0.018
	(0.012)	(0.012)	(0.013)	(0.015)	(0.012)	(0.013)	(0.013)	(0.014)
National Capital		-0.099***	-0.077***	-0.073***		-0.030*	-0.026*	-0.024
		(0.011)	(0.010)	(0.011)		(0.013)	(0.013)	(0.014)
Regional Capital		-0.028**	-0.008	-0.004		0.020	0.024*	0.024*
		(0.010)	(0.010)	(0.010)		(0.011)	(0.011)	(0.011)
DHS Urban			-0.084***	-0.072***			-0.013	-0.016
			(0.009)	(0.011)			(0.010)	(0.011)
ln(pop density)				-0.013***				-0.001
				(0.003)				(0.003)
Region	SSA	SSA	SSA	SSA	SSA	SSA	SSA	SSA
Outcome mean	0.485	0.485	0.485	0.479	0.296	0.296	0.296	0.293
R^2	0.144	0.146	0.149	0.151	0.066	0.067	0.067	0.069
N	397168	397168	397168	362102	115517	115517	115517	102675
Surveys	29	29	29	29	22	22	22	22

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

6 Conclusions

This study examines demographic and health outcomes across cities, towns and rural areas defined consistently at a global scale. To conclude, we highlight our main findings and point out some promising questions for further research.

First, for most outcomes we can measure, related to infrastructure, health, education, fertility and domestic violence, for all regions we study, city residents are on average better off than their rural counterparts. Differentials are larger in utilities and education than in the other categories. The main exceptions to these broad patterns are health conditions most likely to be related to pollution and urban diet, as well as some types of crime, where city residents are at some disadvantage. At the level of individual countries, differentials are less precisely estimated, and there is more variation, with a few countries seeing an urban disadvantage in indicators such as infant mortality, diarrhoea, and contraception use.

Second, town residents are typically in between city and rural, except in South Asia, where they are the worst off in several attributes, perhaps indicating that the fast expansion of towns has outpaced the supply of public services, at least relative to improvements in rural areas. This suggests policy interventions targeting towns may improve development outcomes in the region.

Third, sorting by age, gender, and education can typically explain some but rarely all of these differentials. Alternatively, controlling for national or regional capital status, population density, and administrative status can explain more, but still not all of the city-rural differential.

Fourth, most outcomes have improved noticeably since 2000 in Africa and South Asia and urban-rural differences for some outcomes have shrank over time quite substantially, especially in South Asia. The change is particularly large for infant mortality, which fell by 50% in all zones in South Asia, and in rural Africa. More expensive and less mobile public goods with economies of scale seem to improve more slowly, and are especially concentrated in urban areas, while cheaper and more mobile ones improve faster. For example, the penetration rate of piped water is very low and has improved slowly if at all. In contrast, the spread of relative cheaper public goods such as vaccines and condoms has been quite rapid. Through some combination of these improvements and other sources, health outcomes have shown substantial improvement. Whether budget allocations could be reoptimized between these cheaper and more expensive strategies remains an open question.

We close with a few further thoughts. The mobility of some public goods may be rooted in the preferences of service providers. For example, education may be less mobile if teachers do not want live in rural areas. The ability of the governments to improve the mobility of public services may explain why the urban-rural gap in education expanded in Africa but shrank in South Asia.

While urban advantage persists in Africa, migration plays a role that is relatively unexplored. We note a few findings here but it was not a focus of the study. One is that the penetration rate of piped water is reduced from 2000 to 2015 in urban Africa, which may be partially driven by migrants on the edge of cities. Another is that residents in African national capitals have some worse outcomes than other city residents, where these cities are a particular draw for migrants. Finally, recent literature argues that frictions in migration due to travel costs in poor countries may impair the ability of an economy to adjust towards a spatial equilibrium where urban and rural residents share the same level of utility (Chauvin et al., 2017; Gollin et al., 2018). Transport investments play a role in those frictions (Baum-Snow et al., 2017; Jedwab and Storeygard, 2017) and may in part explain why rural residents stay in places with low incomes and services.

We provide suggestive evidence that political favoritism, agglomeration economies through higher density, and urban administrative status might explain a great deal of the city effect. For example, access to utilities are 13% to 15% higher in national capitals than non-capital cities, and 20% to 30% higher in administratively (DHS) urban than DHS rural areas. However, even setting aside identification issues, there are still residual city effects for most outcomes even after controlling for these factors. It is possible that remaining city effects could also stem from aspects of urban culture and history that we have not measured. Future research should investigate them further.

A Demographic and Health Surveys

Table A.1: Demographic and Health Surveys

		Survey year		c. 2015 sample size					c. 2000 sample size				
Region	Country	c. 2015	c. 2000	Female	Male	HH	Age 16	Under 5	Female	Male	HH	Age 16	Under 5
SSA	Angola	2015-16		14290	5654	16031	1370	14259					
SSA	Benin	2011-12	2001	16358	5112	17174	1479	13212	6137	2658	5700	543	5311
SSA	Burkina Faso	2010	1998-99	16072	6886	13566	1357	14106	6379	2612	4767	631	5884
SSA	Burundi	2010		9372	4270	8574	936	7721					
SSA	DR Congo	2013-14		17273	7919	16544	1656	17053					
SSA	Cameroon	2011		15286	7118	14084	1403	11585					
SSA	Chad	2014-15		17696	5236	17208	1822	18604					
SSA	Comoros	2012		4816	1960	4057	520	2799					
SSA	Cote d'Ivoire	2011-12	1998-99	9735	4976	9336	839	7496	3029	883	2113	302	1980
SSA	Ethiopia	2016	2000	15242	12367	16157	1448	10221	15215	2575	13942	1610	10786
SSA	Gabon	2012		8299	5536	9583	827	5989					
SSA	Ghana	2014	1998	9222	4306	11607	806	5772	4755	1516	5883	416	3235
SSA	Guinea	2012	1999	9114	3774	7087	775	7023	6609	1925	4992	605	5748
SSA	Kenya	2014	2003	30776	12674	36024	3387	20742	8035	3504	8406	880	5834
SSA	Lesotho	2014		6580	2919	9356	978	3124					
SSA	Liberia	2013		9102	4055	9214	868	7520					
SSA	Malawi	2015-16	2000	24202	7368	26022	2268	17050	12926	3029	13885	1312	11637
SSA	Mali	2012-13	2001	10407	4395	10087	901	10309	12680	3375	12143	1198	12918

Table A.1: Demographic and Health Surveys

SSA	Mozambique	2011		13636	4006	13817	1222	11024					
SSA	Namibia	2013	2000	10007	4475	9832	830	5035	6625	2913	6267	659	3941
SSA	Nigeria	2013	2003	38491	17163	38080	2956	31124	7546	2312	7145	652	5965
SSA	Rwanda	2014-15		13332	6139	12545	1055	7752					
SSA	Senegal	2010-11	1997	15409	4827	7740	1429	12054	8465	4248	4701	957	7238
SSA	Sierra Leone	2013		16489	7192	12489	1294	11822					
SSA	Tanzania	2015-16	1999	13204	3497	12503	1288	10181	3872	3414	3488	402	3090
SSA	Togo	2013-14	1998	9480	4476	9549	775	6979	8520	3802	7460	804	4148
SSA	Uganda	2016	2000-01	18013	5204	19024	2032	15094	6161	1656	6701	717	5961
SSA	Zambia	2013-14		16305	14678	15805	1740	13342					
SSA	Zimbabwe	2015	1999	9885	8350	10480	937	6092	5721	2527	6156	812	3543
S Asia	Bangladesh	2014	1999-00	17456		16922	1791	7711	10489	2543	9804	1383	6781
S Asia	India	2015-16		654524	104864	559466	52528	243777					
S Asia	Nepal	2016	2001	12820	4047	11011	1064	5027	8544	2219	8436	992	6791
LAC	Colombia	2010		51518		49595	4108	17133					
LAC	Dominican Republic	2013		9218	10129	11287	832	3666					
LAC	Guatemala	2014-15		25700	11055	21211	2460	12313					
LAC	Haiti	2012		13843	9210	12795	1440	7063					
LAC	Honduras	2011-12		22121	6942	20794	2484	10596					
SE Asia	Cambodia	2014		17469	5155	15720	1284	7111					
SE Asia	Myanmar	2015-16		12785	4692	12389	922	4783					
SE Asia	Timor-Leste	2016		12549	4600	11450	1364	7186					

B Heterogeneity Across Countries c. 2015

Figure B.1: Electricity

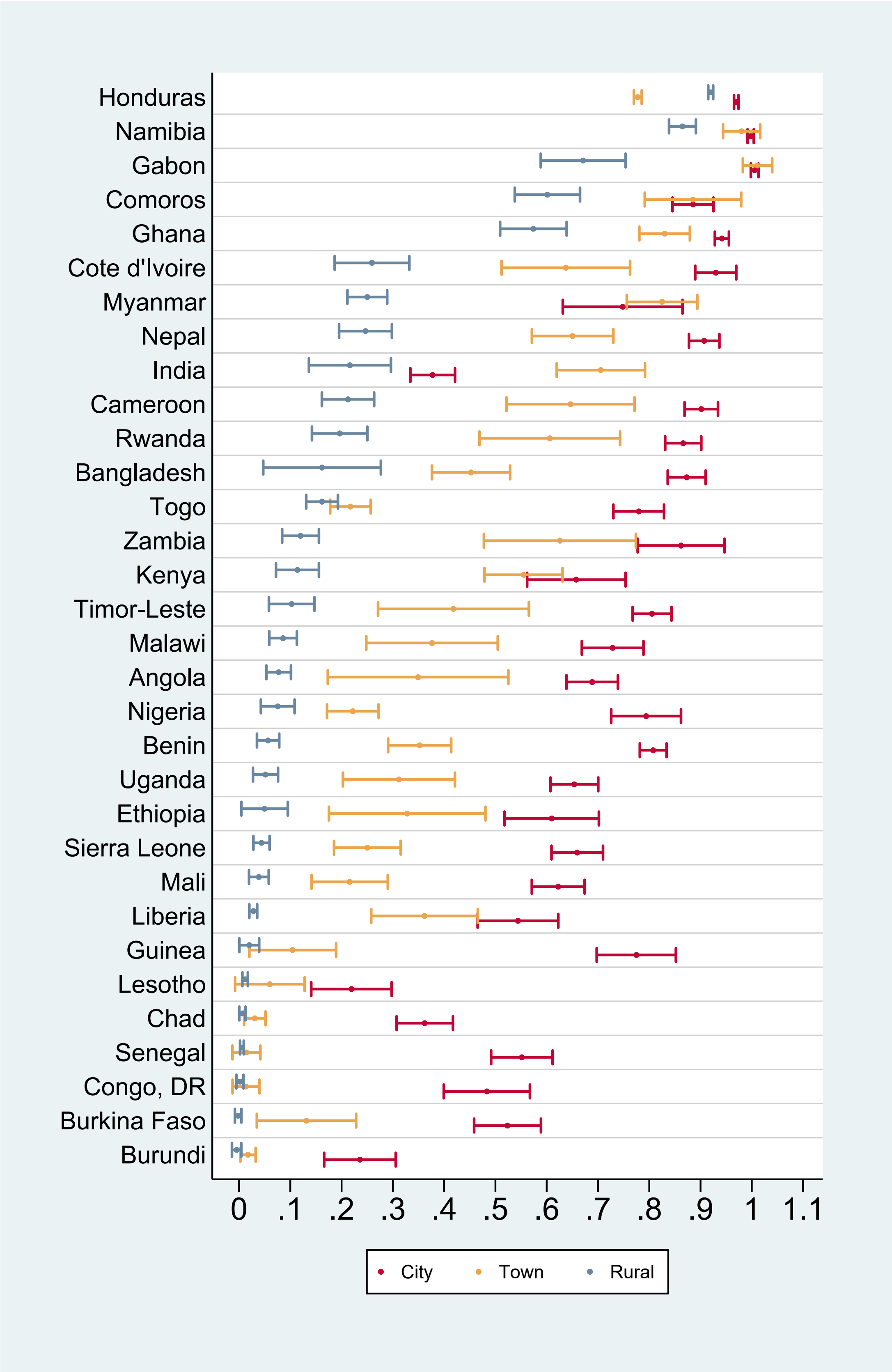
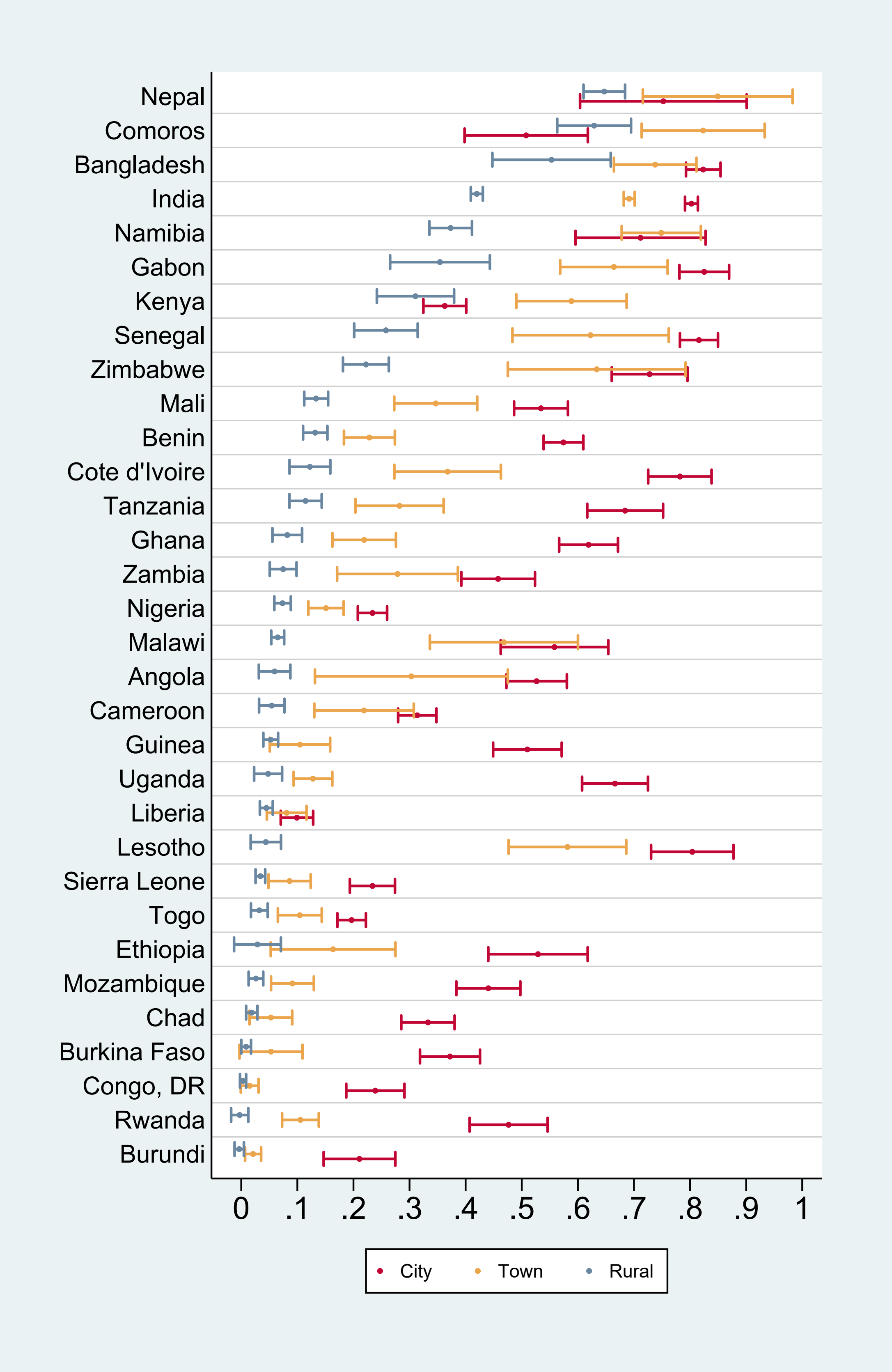
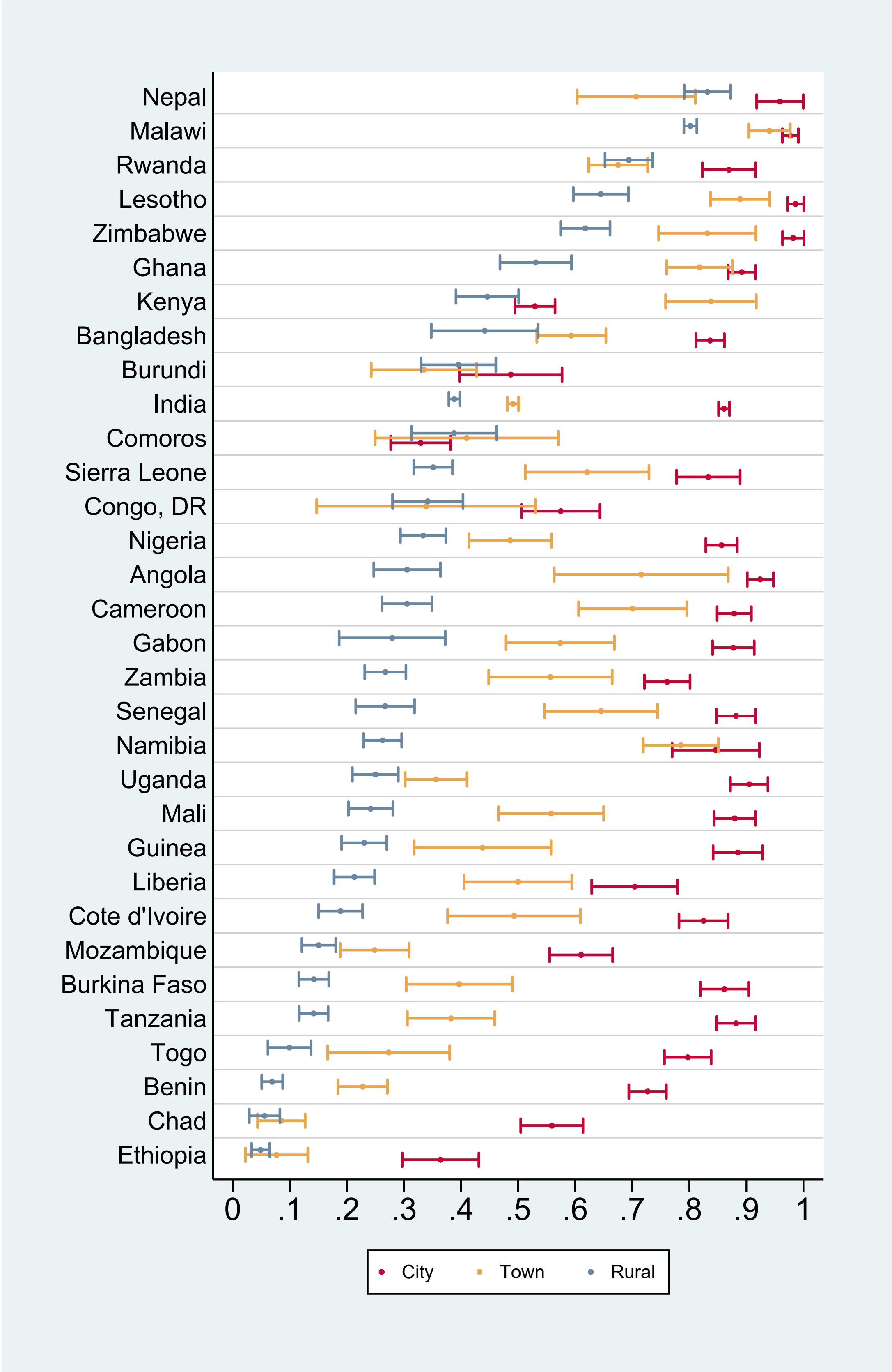


Figure B.2: Safely managed drinking water



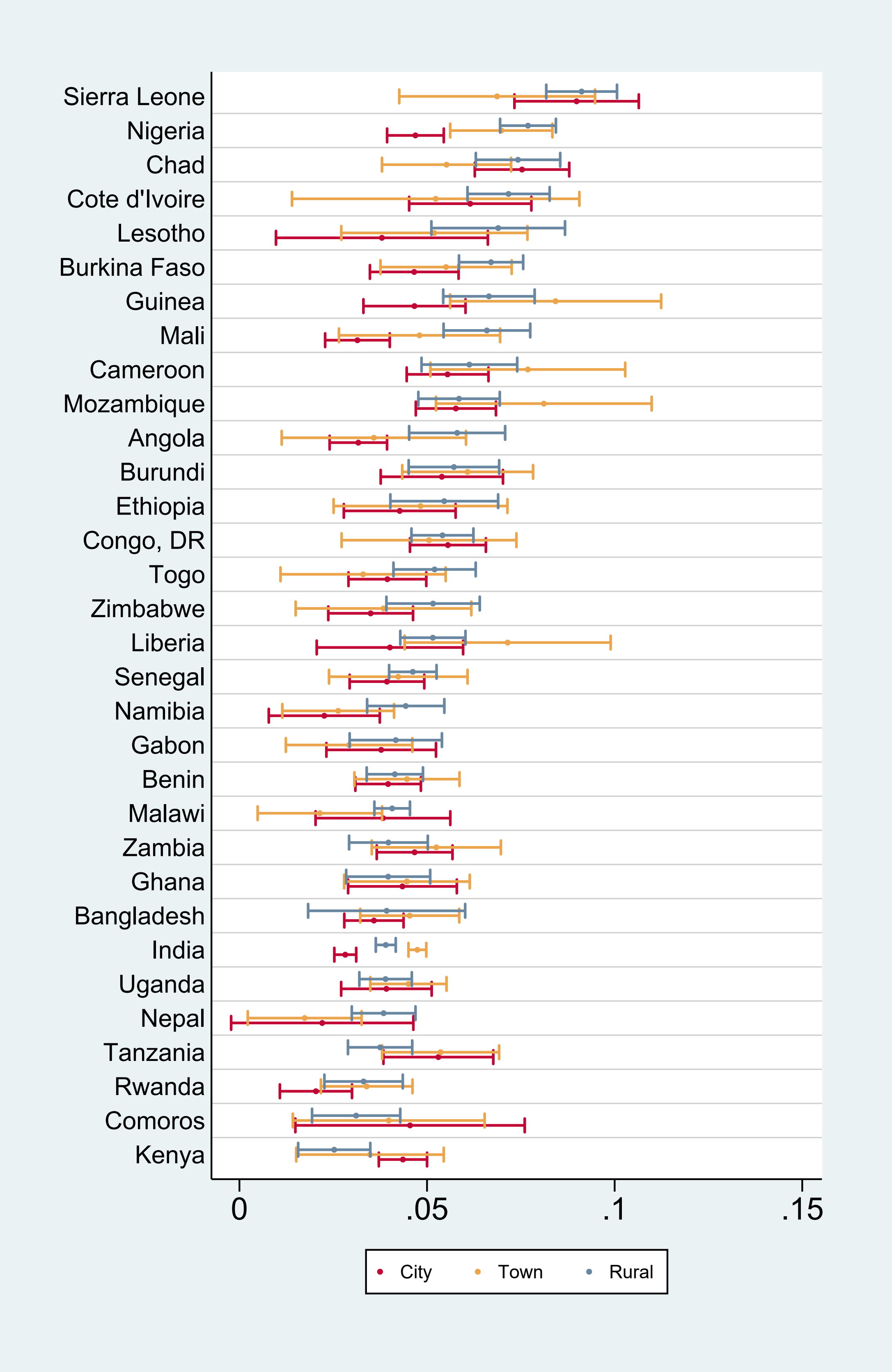
Safely managed drinking water is defined by the DHS-WHO Joint Monitoring Program as all improved water sources that take zero minutes to collect, or are on the premises. Improved water sources encompass all piped water and packaged water, as well as protected wells or springs, boreholes, and rainwater.

Figure B.3: Improved sanitation



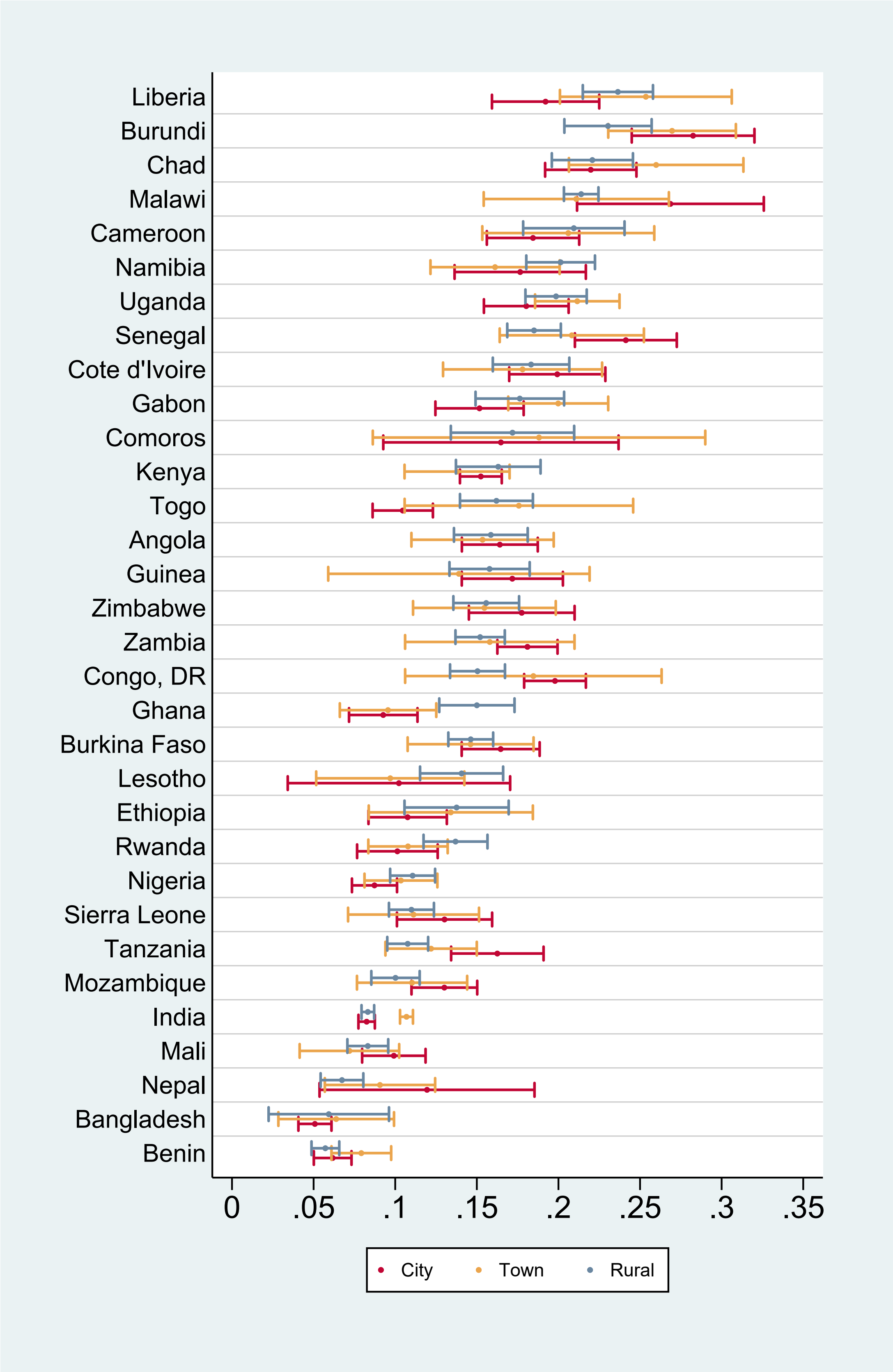
Improved sanitation is defined by the DHS-WHO Joint Monitory Program to include: all shared and non-shared facilities that flush/pour flush to piped sewer system, septic tank, pit latrine; ventilated improved pit latrine, pit latrine with slab, and composting toilet. Additionally, facilities that flush to unknown locations are considered improved, whereas facilities that flush to a known location but not to a sewer system, septic tank, or pit latrine are classified as unimproved.

Figure B.4: Infant death



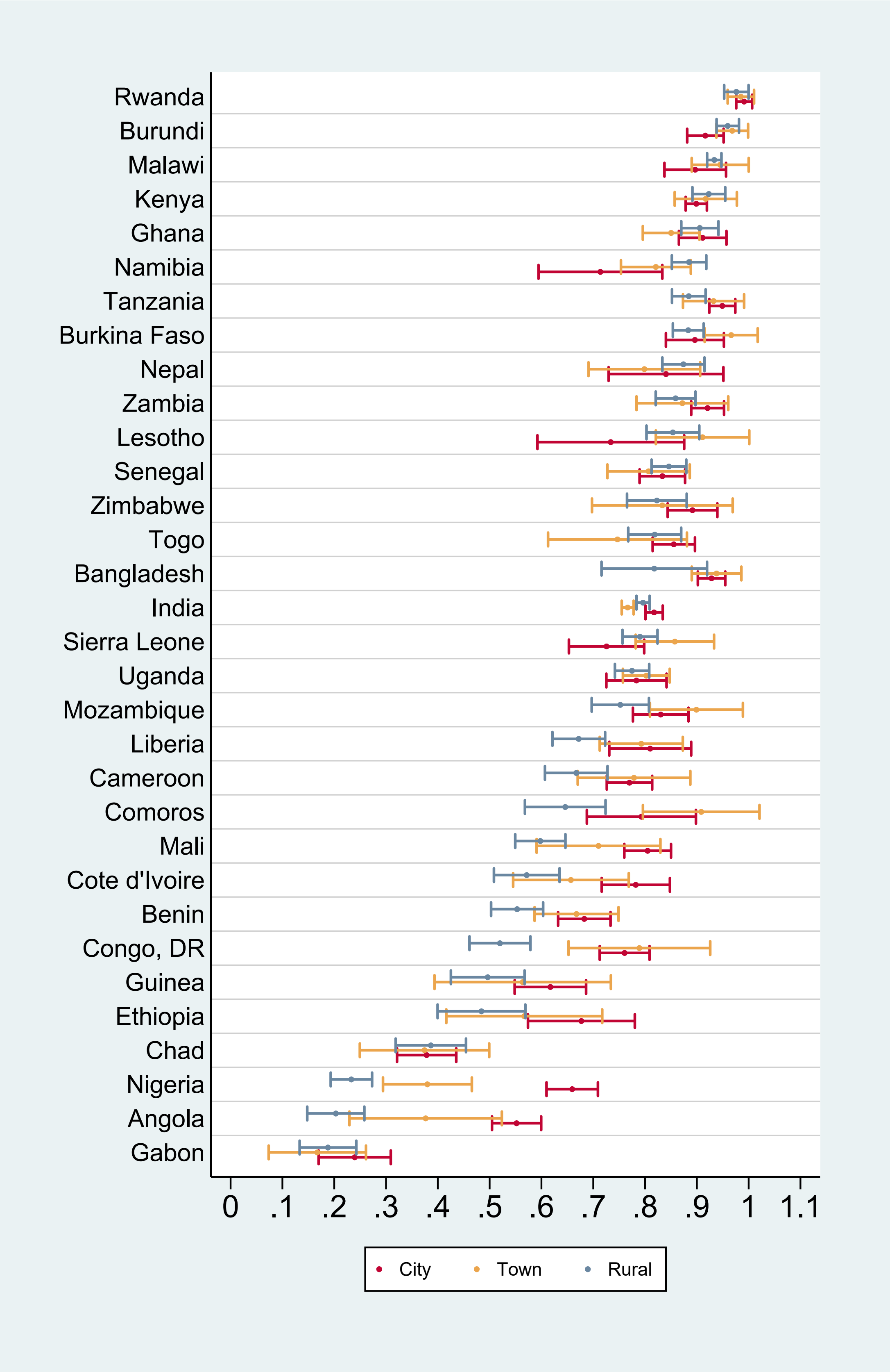
The dependent variable infant death is a dummy for whether or not the child died within 1 year of being born, for children who were born between 1 and 5 years before the survey.

Figure B.5: Diarrhoea



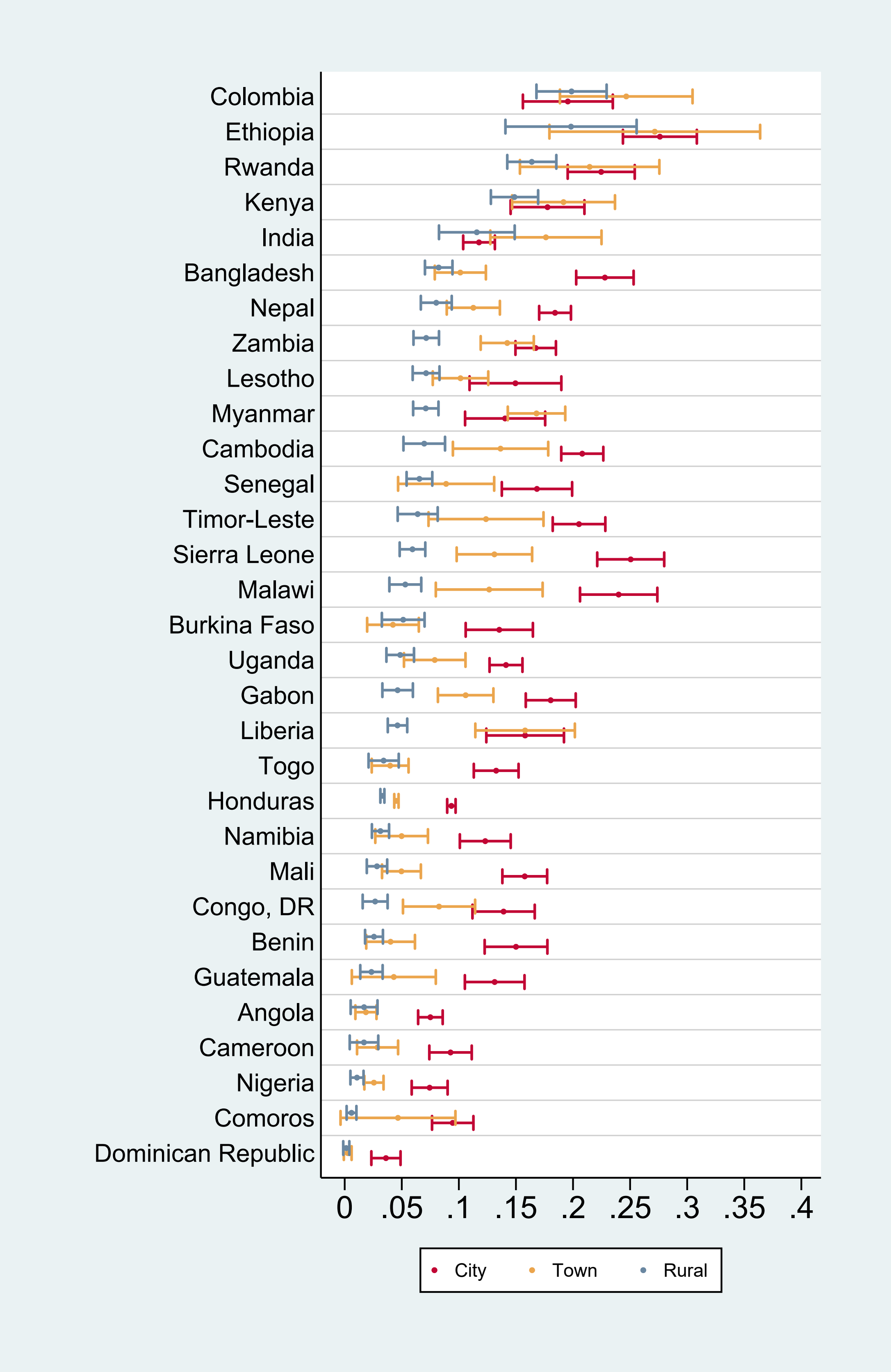
Each point is the share of children aged 5 and under who have had diarrheoa in the last two weeks.

Figure B.6: Third dose of DPT vaccine



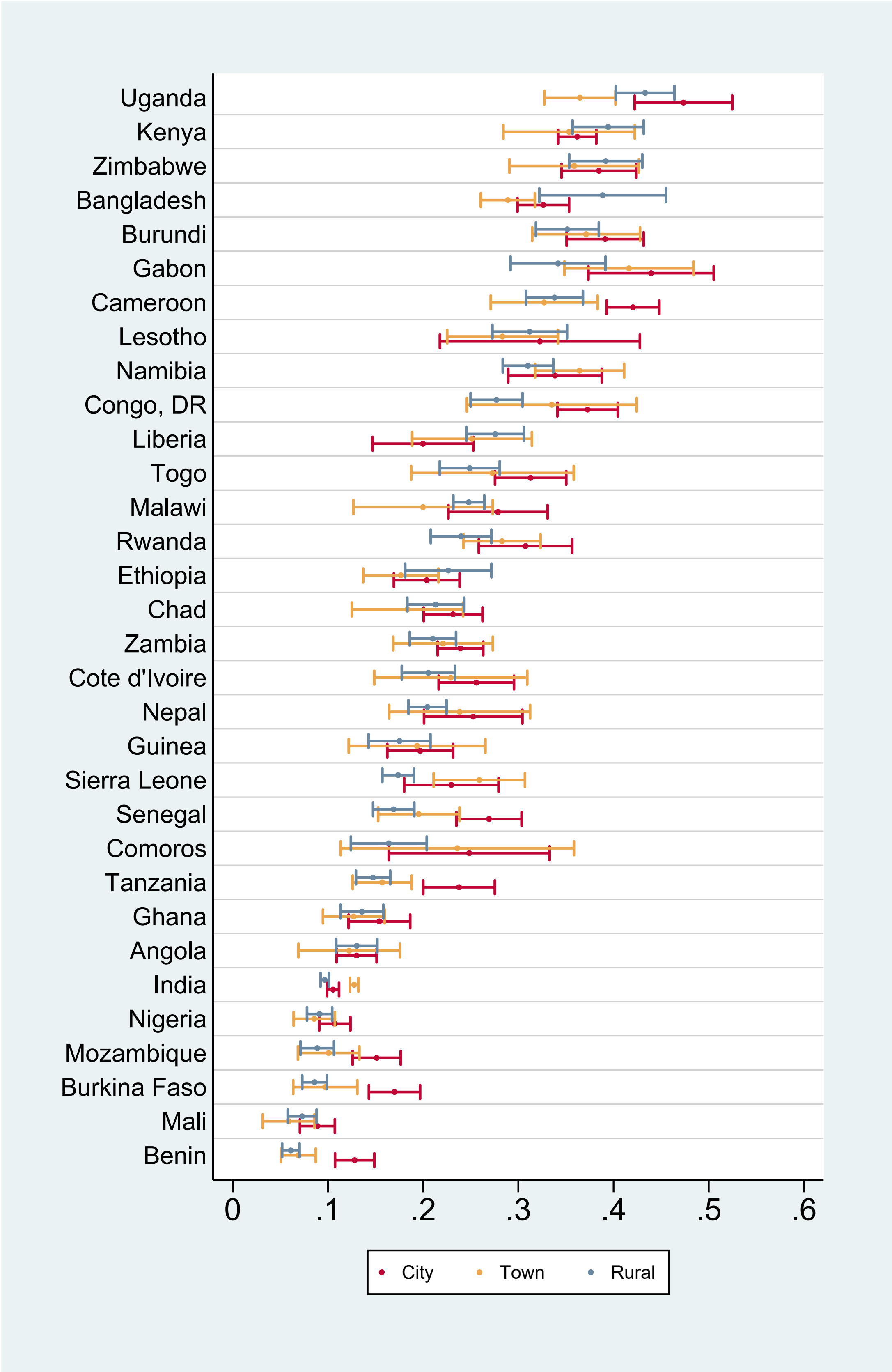
Each point is the share of children aged 12-24 months that received the third dose of the DPT vaccine.

Figure B.7: Adults with obesity



Obesity is defined as having a body mass index ≥ 30 for 20–49-year-olds. The sample consists of household members that were 20–49 years old, eligible for the individual male and female surveys, and not pregnant.

Figure B.8: Cough



Each point is the share of children aged 5 and under who have had a cough in the last two weeks.

Figure B.9: 8 years of schooling among 16 year olds

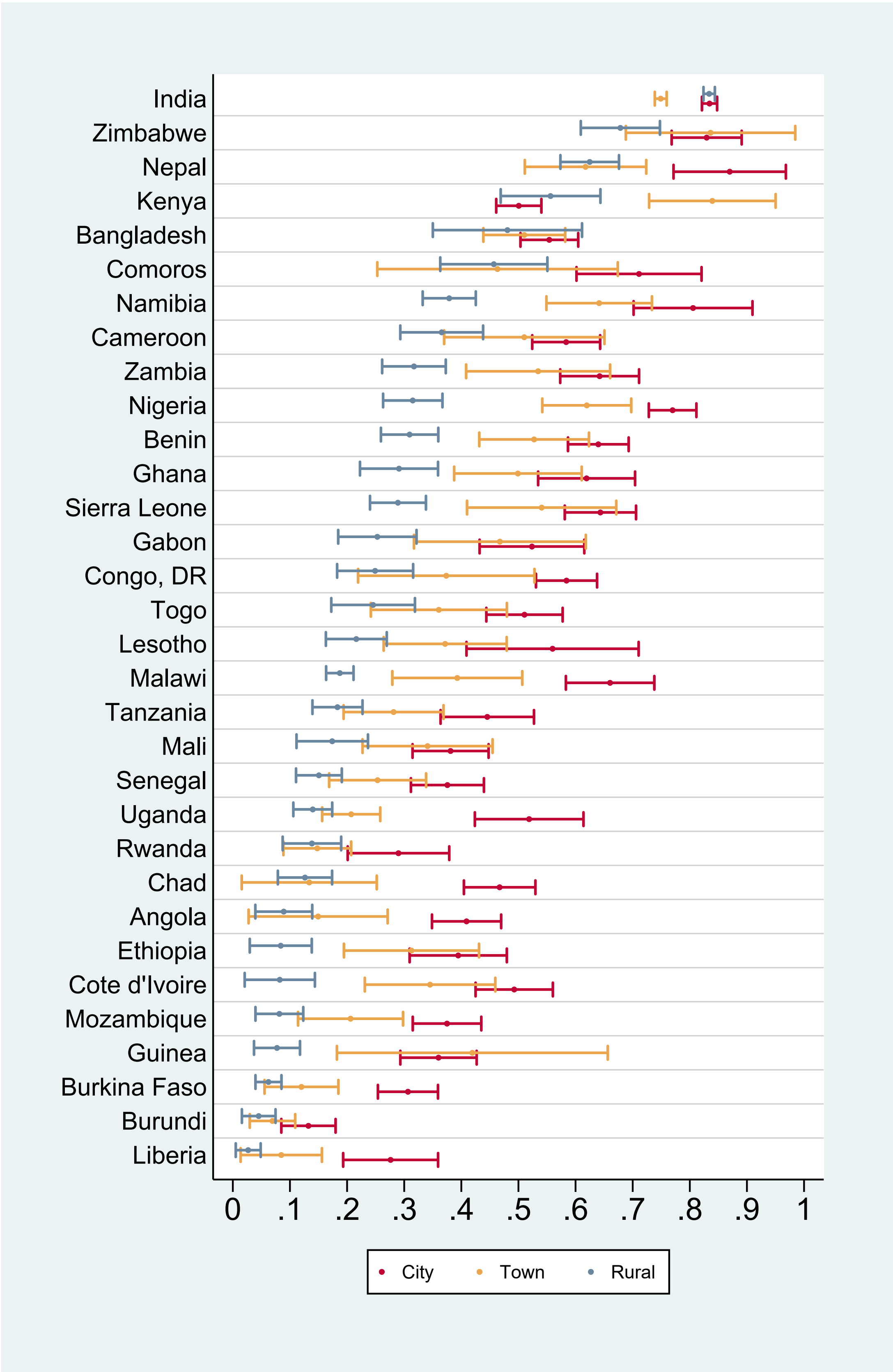
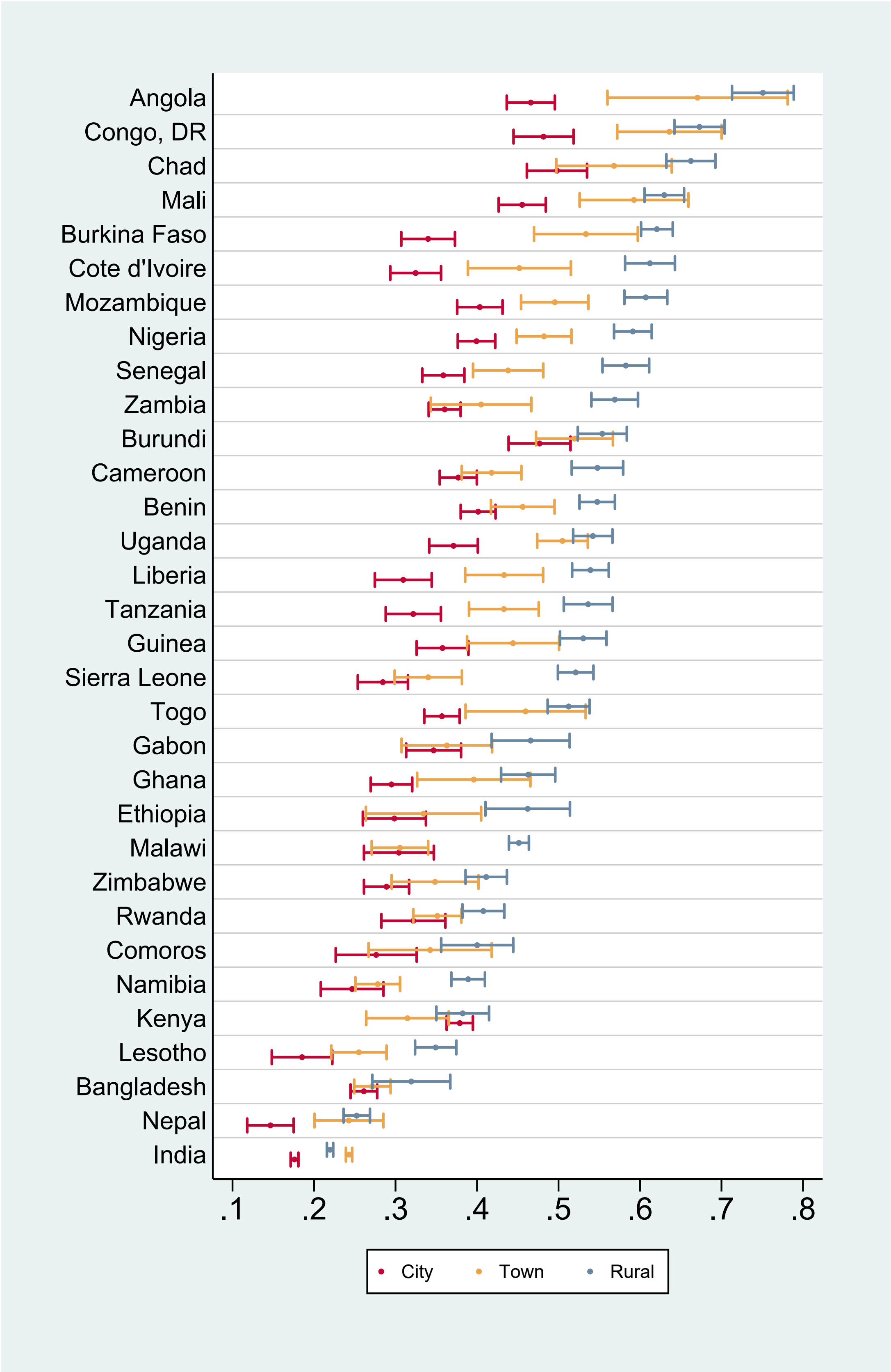
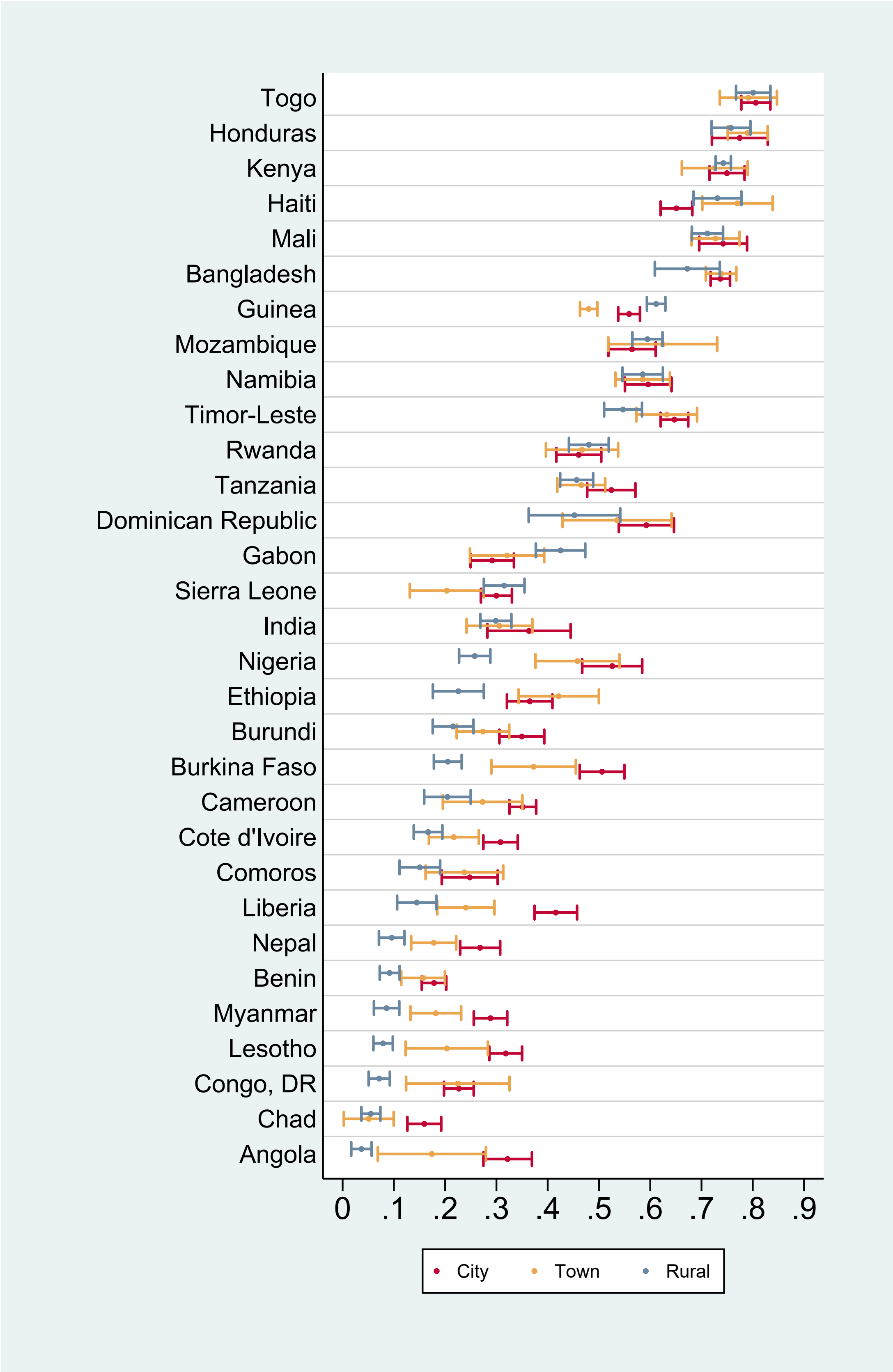


Figure B.10: Total number of births per woman in three years



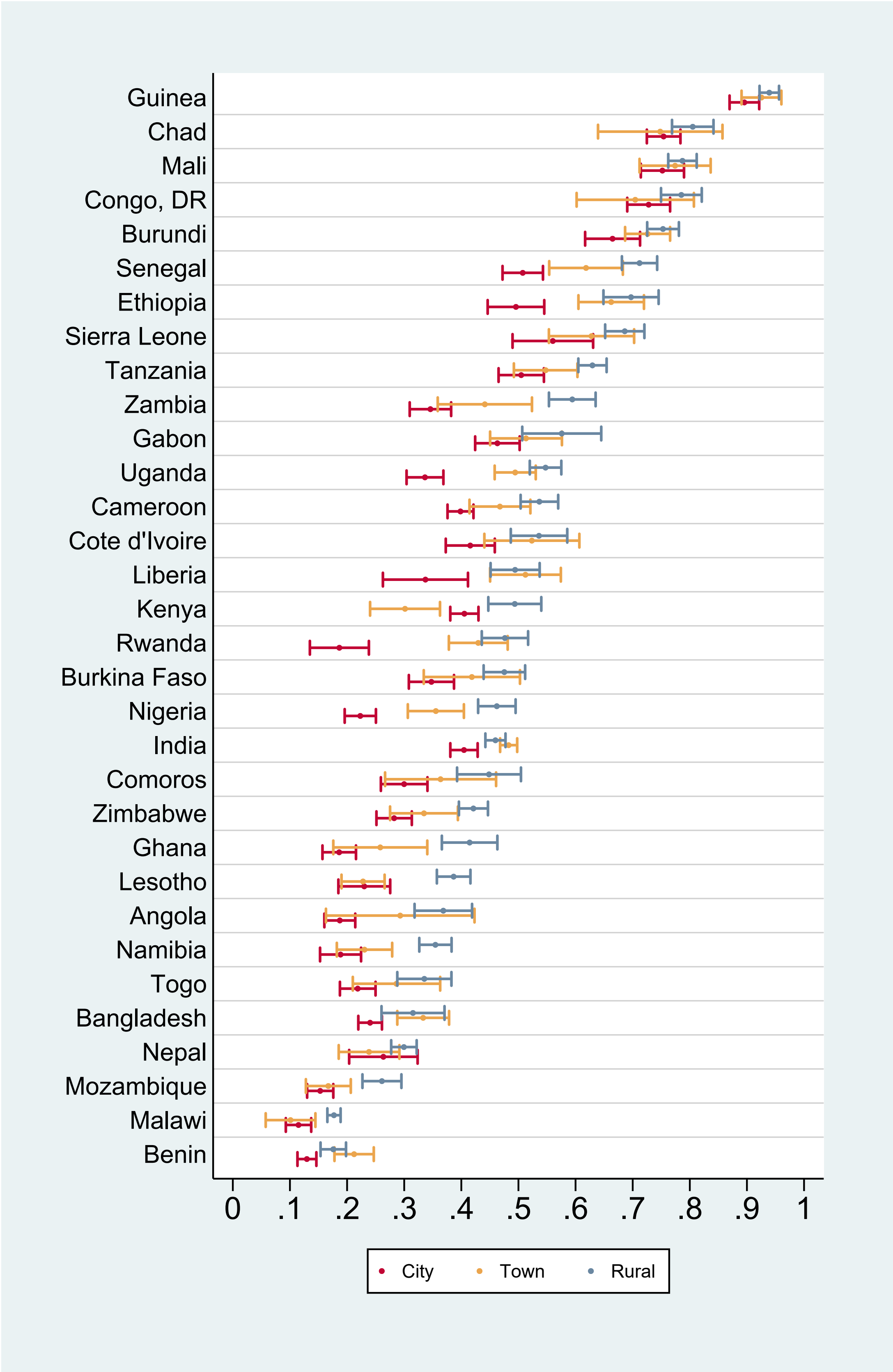
Each point is the total number of births per woman in the past three years for women aged 20–40.

Figure B.11: Current use of modern contraception



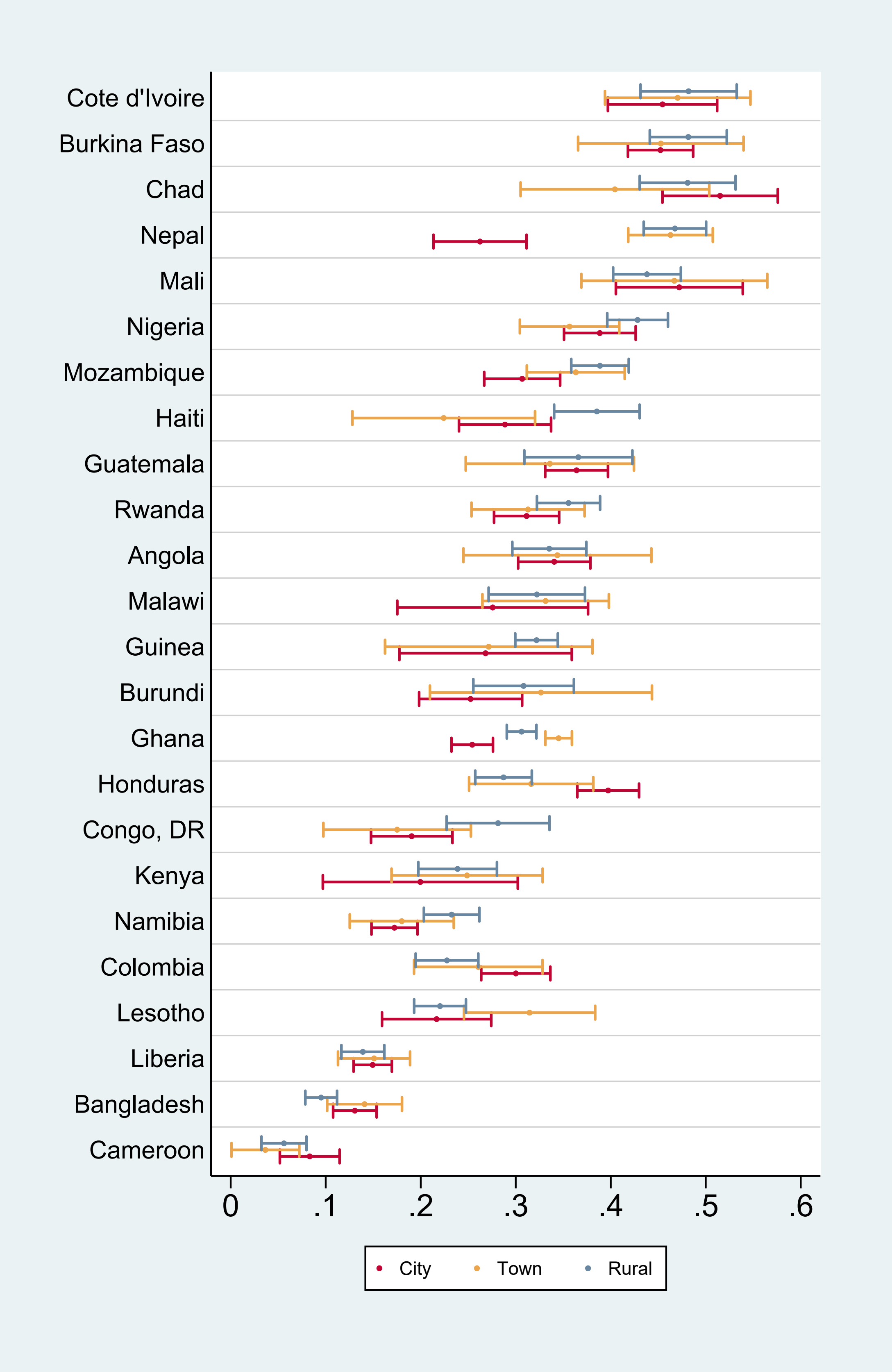
Each point is the share of women aged 20–40 currently using modern contraception. Modern contraception includes: pill, IUD, injections, diaphragm, condom, female sterilisation, male sterilisation, implants, female condom, foam and jelly, emergency contraception, other modern.

Figure B.12: Women who think wife beatings are justified



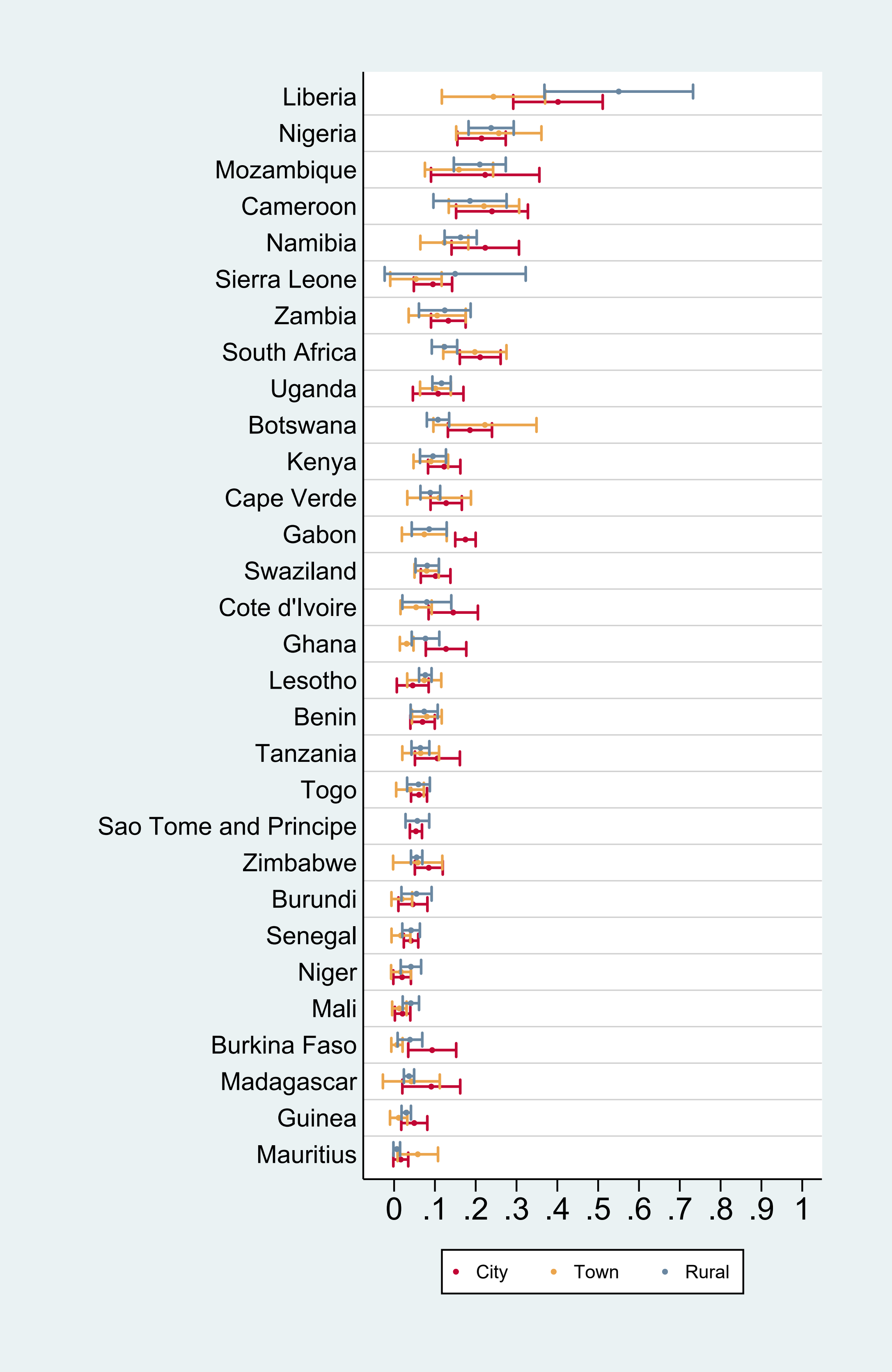
Each point is the share of women aged 15–49 whose opinion is that a husband is justified in hitting or beating his wife when: she goes out without telling him, she argues with him, she refuses to have sex with him, or she burns the food.

Figure B.13: Share of women who have ever experienced domestic violence



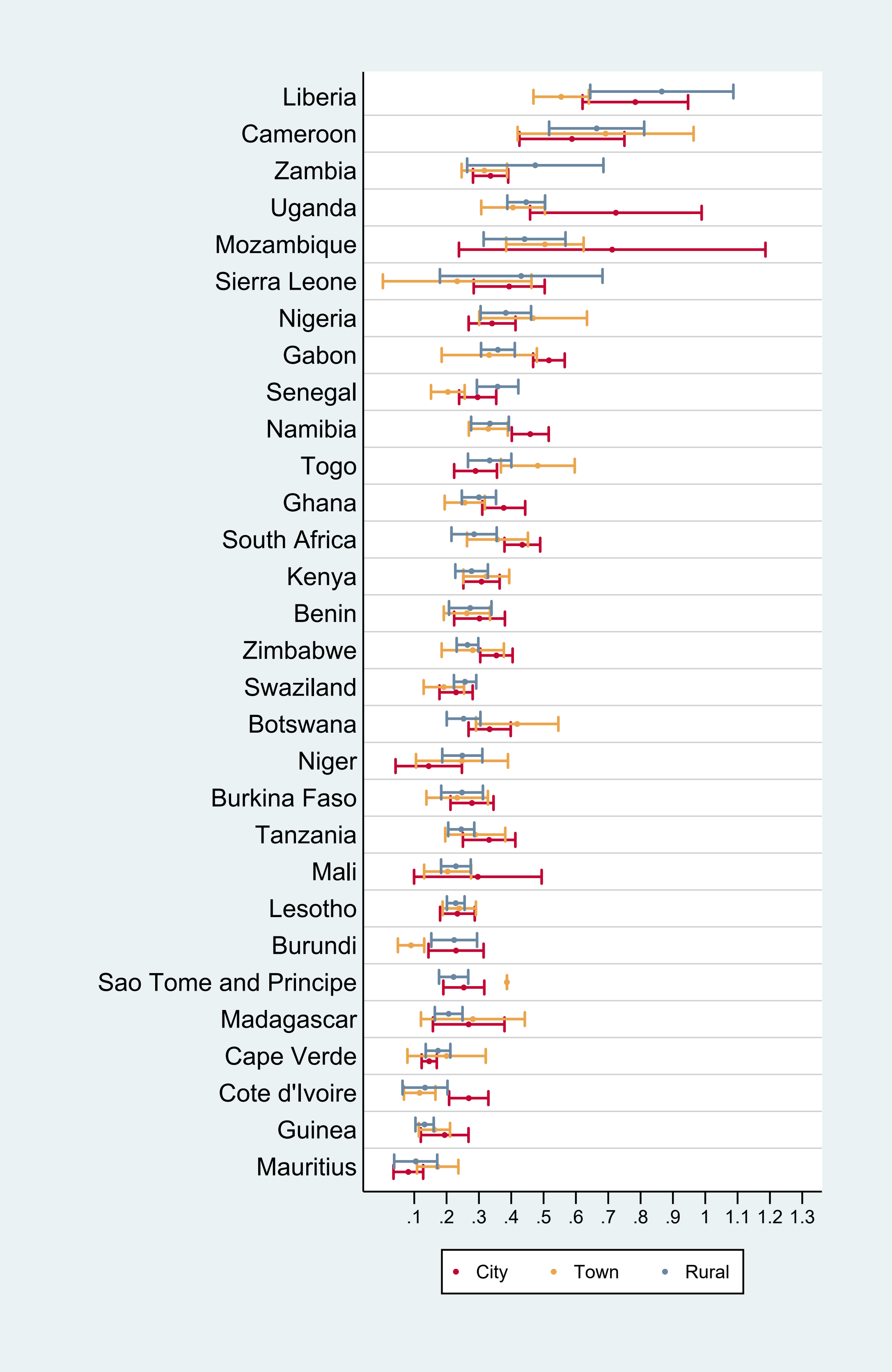
Each point is the share of currently married women aged 15–49 who have ever experienced domestic violence from their current spouse.

Figure B.14: Been physically attacked



Each point is the share of households who responded that a member of the family had been physically attacked in the past year.

Figure B.15: Had something stolen from house



Each point is the share of households with family members that had something stolen from their house in the past year

C Sorting regression tables

Table C.1: Households with electricity

	(1)	(2)	(3)	(4)
City	0.583***	0.485***	0.096***	0.061***
	(0.011)	(0.011)	(0.006)	(0.006)
Town	0.279***	0.235***	-0.135***	-0.142***
	(0.019)	(0.017)	(0.007)	(0.007)
ln(hh size)		-0.015***		0.012***
		(0.002)		(0.002)
Female household head		0.036***		0.014***
		(0.004)		(0.003)
ln(age hh head)		0.036***		0.088***
		(0.005)		(0.004)
Education of hh head: Primary		0.093***		0.089***
		(0.006)		(0.003)
Education of hh head: Secondary		0.241***		0.147***
		(0.008)		(0.003)
Education of hh head: Higher		0.394***		0.186***
		(0.010)		(0.004)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.357	0.357	0.849	0.849
R^2	0.391	0.441	0.096	0.131
N	413359	409493	587399	584553
Surveys	29	29	3	3

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.2: Households with safely managed drinking water

	(1)	(2)	(3)	(4)
City	0.359***	0.298***	0.345***	0.300***
	(0.009)	(0.009)	(0.008)	(0.008)
Town	0.133***	0.107***	0.235***	0.223***
	(0.011)	(0.010)	(0.009)	(0.009)
ln(hh size)		-0.015***		0.053***
		(0.002)		(0.002)
Female household head		0.022***		0.051***
		(0.003)		(0.003)
ln(age hh head)		0.035***		0.122***
		(0.004)		(0.004)
Education of hh head: Primary		0.030***		0.040***
		(0.004)		(0.003)
Education of hh head: Secondary		0.116***		0.140***
		(0.005)		(0.003)
Education of hh head: Higher		0.270***		0.250***
		(0.009)		(0.005)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.204	0.203	0.647	0.648
R^2	0.207	0.236	0.062	0.092
N	413575	409707	587399	584553
Surveys	29	29	3	3

Safely managed drinking water is defined by the DHS-WHO Joint Monitoring Program as all improved water sources that take zero minutes to collect, or are on the premises. Improved water sources encompass all piped water and packaged water, as well as protected wells or springs, boreholes, and rainwater. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.3: Infant death

	(1)	(2)	(3)	(4)
City	-0.0094***	-0.0040	-0.0094***	-0.0045*
	(0.0021)	(0.0022)	(0.0019)	(0.0020)
Town	0.0005	0.0023	0.0084***	0.0081***
	(0.0033)	(0.0033)	(0.0020)	(0.0020)
ln(age of child)		0.0047**		-0.0015
		(0.0016)		(0.0014)
Female child		-0.0109***		-0.0061***
		(0.0015)		(0.0013)
Mother's age at birth/10,		-0.1582***		-0.0808
		(0.0445)		(0.0811)
Mother's age at birth/10, squared		0.0432**		0.0108
		(0.0160)		(0.0309)
Mother's age at birth/10, cubed		-0.0035		0.0011
		(0.0019)		(0.0038)
Education of mom: Primary		-0.0031		-0.0045*
		(0.0025)		(0.0021)
Education of mom: Secondary		-0.0095**		-0.0137***
		(0.0030)		(0.0018)
Education of mom: Higher		-0.0111*		-0.0217***
		(0.0050)		(0.0024)
Female household head		-0.0014		-0.0042*
		(0.0020)		(0.0018)
ln(age of hh head)		-0.0017		-0.0037
		(0.0030)		(0.0022)
Education of hh head: Primary		0.0016		-0.0011
		(0.0026)		(0.0022)
Education of hh head: Secondary		-0.0029		-0.0032
		(0.0030)		(0.0017)
Education of hh head: Higher		-0.0023		-0.0111***
		(0.0050)		(0.0024)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.053	0.053	0.040	0.040
R^2	0.003	0.005	0.001	0.005
N	257671	254877	206657	205092
Surveys	29	29	3	3

The dependent variable is a dummy for whether or not a child died within 1 year of being born for children who were born between 1 and 5 years before the survey. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.4: Diarrhoea in past two weeks

	(1)	(2)	(3)	(4)
City	0.000	0.010*	0.001	0.004
	(0.004)	(0.004)	(0.003)	(0.003)
Town	0.002	0.006	0.023***	0.023***
	(0.007)	(0.007)	(0.003)	(0.003)
ln(child age)		-0.027***		-0.032***
		(0.001)		(0.001)
Female child		-0.011***		-0.005**
		(0.002)		(0.002)
Mother's age at birth/10,		0.068		0.097
		(0.060)		(0.067)
Mother's age at birth/10, squared		-0.037		-0.033
		(0.021)		(0.024)
Mother's age at birth/10, cubed		0.005*		0.004
		(0.002)		(0.003)
Education of mom: Primary		0.003		0.003
		(0.004)		(0.003)
Education of mom: Secondary		-0.009*		-0.005*
		(0.004)		(0.002)
Education of mom: Higher		-0.046***		-0.014***
		(0.007)		(0.004)
Female household head		-0.002		0.006*
		(0.003)		(0.003)
ln(age hh head)		-0.003		-0.009***
		(0.004)		(0.003)
Education of hh head: Primary		0.001		-0.000
		(0.004)		(0.002)
Education of hh head: Secondary		-0.008		-0.002
		(0.004)		(0.002)
Education of hh head: Higher		-0.013*		-0.004
		(0.006)		(0.004)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.142	0.144	0.089	0.090
R^2	0.014	0.021	0.002	0.012
N	297759	291201	244212	240551
Surveys	29	29	3	3

The dependent variable is a dummy for whether a child aged 5 and under had diarrhoea in the last two weeks. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.5: DPT 3 Vaccination

	(1)	(2)	(3)	(4)
City	0.211***	0.099***	0.026**	0.005
	(0.014)	(0.013)	(0.010)	(0.010)
Town	0.109***	0.059***	-0.017	-0.010
	(0.019)	(0.017)	(0.010)	(0.009)
ln(child age)		0.021		0.045**
		(0.017)		(0.015)
Female child		-0.005		0.004
		(0.007)		(0.005)
Mother's age at birth/10,		0.400*		0.189
		(0.181)		(0.211)
Mother's age at birth/10, squared		-0.104		-0.062
		(0.065)		(0.078)
Mother's age at birth/10, cubed		0.009		0.005
		(0.007)		(0.009)
Education of mom: Primary		0.111***		0.077***
		(0.010)		(0.009)
Education of mom: Secondary		0.246***		0.117***
		(0.013)		(0.008)
Education of mom: Higher		0.323***		0.143***
		(0.022)		(0.011)
Female household head		0.015		0.025**
		(0.008)		(0.008)
ln(age hh head)		0.013		0.023**
		(0.012)		(0.009)
Education of hh head: Primary		0.073***		0.030***
		(0.011)		(0.008)
Education of hh head: Secondary		0.081***		0.041***
		(0.012)		(0.007)
Education of hh head: Higher		0.096***		0.044***
		(0.019)		(0.012)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.642	0.642	0.802	0.802
R^2	0.211	0.262	0.012	0.038
N	55768	55130	44770	44416
Surveys	29	29	3	3

The dependent variable is a dummy for whether a child aged 12-24 months received the third dose of the DPT vaccination. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.6: Adults with obesity

	(1)	(2)	(3)	(4)
City	0.100***	0.076***	0.060***	0.048***
	(0.003)	(0.003)	(0.002)	(0.002)
Town	0.035***	0.023***	0.010***	0.007***
	(0.005)	(0.005)	(0.001)	(0.001)
Respondent's age/10		-0.073		-0.020
		(0.056)		(0.030)
Respondent's age/10, squared		0.051**		0.024*
		(0.017)		(0.009)
Respondent's age/10, cubed		-0.006***		-0.003**
		(0.002)		(0.001)
Female respondent		0.057***		0.032***
		(0.003)		(0.001)
Education of respondent: Primary		0.023***		0.019***
		(0.002)		(0.001)
Education of respondent: Secondary		0.040***		0.037***
		(0.004)		(0.001)
Education of respondent: Higher		0.052***		0.032***
		(0.006)		(0.002)
ln(age of hh head)		0.012***		0.011***
		(0.003)		(0.002)
Female household head		0.006**		0.002
		(0.002)		(0.001)
Education of hh head: Primary		0.005		0.003*
		(0.003)		(0.001)
Education of hh head: Secondary		0.023***		0.014***
		(0.004)		(0.001)
Education of hh head: Higher		0.045***		0.033***
		(0.006)		(0.002)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.088	0.088	0.053	0.053
R^2	0.049	0.073	0.010	0.031
N	227706	225076	602570	598988
Surveys	28	28	3	3

Obesity is defined as having a body mass index ≥ 30. The sample includes household members eligible for the individual male and female surveys, aged 20–49 and not pregnant. Data on obesity for the Angola 2015–16 survey was unavailable. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for p $\leq$ 0.05, ** for p $\leq$ 0.01, and *** for p $\leq$ 0.001.

Table C.7: Children with cough

	(1)	(2)	(3)	(4)
City	0.035***	0.021***	0.009*	0.009*
	(0.005)	(0.006)	(0.004)	(0.004)
Town	-0.003	-0.010	0.026***	0.027***
	(0.007)	(0.008)	(0.004)	(0.004)
ln(child age)		-0.012***		-0.015***
		(0.001)		(0.001)
Female child		-0.001		-0.011***
		(0.002)		(0.002)
Mother's age at birth/10,		0.115		-0.044
		(0.066)		(0.112)
Mother's age at birth/10, squared		-0.047*		0.011
		(0.023)		(0.041)
Mother's age at birth/10, cubed		0.006*		-0.001
		(0.003)		(0.005)
Education of mom: Primary		0.021***		0.019***
		(0.004)		(0.004)
Education of mom: Secondary		0.037***		0.010**
		(0.005)		(0.003)
Education of mom: Higher		0.037***		0.007
		(0.011)		(0.005)
Female household head		0.009*		0.005
		(0.004)		(0.004)
ln(age of hh head)		0.004		-0.013***
		(0.005)		(0.004)
Education of hh head: Primary		0.017***		0.010**
		(0.004)		(0.004)
Education of hh head: Secondary		0.020***		0.008*
		(0.005)		(0.003)
Education of hh head: Higher		0.000		0.005
		(0.008)		(0.006)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.208	0.210	0.137	0.137
R^2	0.070	0.073	0.037	0.040
N	297866	291314	244312	240649
Surveys	29	29	3	3

The sample includes all children aged 5 and under. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.8: High blood pressure, India 2015–16

	(1)	(2)
City	0.044***	0.035***
	(0.005)	(0.005)
Town	0.022***	0.024***
	(0.005)	(0.004)
Respondent's age/10		-2.243***
		(0.045)
Respondent's age/10, squared		0.517***
		(0.012)
Respondent's age/10, cubed		-0.036***
		(0.001)
Female respondent		0.002
		(0.003)
Education of respondent: Primary		0.026***
		(0.003)
Education of respondent: Secondary		0.047***
		(0.003)
Education of respondent: Higher		0.041***
		(0.004)
ln(age of hh head)		-0.013***
		(0.004)
Female household head		0.007*
		(0.003)
Education of hh head: Primary		0.006*
		(0.003)
Education of hh head: Secondary		0.011***
		(0.003)
Education of hh head: Higher		0.015**
		(0.005)
Country	India	India
Outcome mean	0.245	0.245
R^2	0.001	0.052
N	508690	505634
Surveys	2	2

The sample includes all household members aged 25 and up. High blood pressure is defined as a mean systolic measure $\geq$ 140 or a mean diastolic measure $\geq$ 30 over three blood pressure readings. Respondents who reported taking medication for hypertension, or who reported that a medical professional told them they were hypertensive, were also coded as having high blood pressure. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for p $\leq$ 0.05, ** for p $\leq$ 0.01, and *** for p $\leq$ 0.001.

Table C.9: Asthma, India 2015–16

	(1)	(2)
City	0.003***	0.004***
	(0.001)	(0.001)
Town	0.003***	0.004***
	(0.001)	(0.001)
Respondent's age/10		-0.021*
		(0.009)
Respondent's age/10, squared		0.008**
		(0.003)
Respondent's age/10, cubed		-0.001
		(0.000)
Female respondent		0.007***
		(0.001)
Education of respondent: Primary		0.004***
		(0.001)
Education of respondent: Secondary		0.003***
		(0.001)
Education of respondent: Higher		0.000
		(0.001)
ln(age of hh head)		-0.002*
		(0.001)
Female household head		0.000
		(0.001)
Education of hh head: Primary		0.001
		(0.001)
Education of hh head: Secondary		-0.000
		(0.001)
Education of hh head: Higher		-0.004***
		(0.001)
Country	India	India
Outcome mean	0.018	0.018
R^2	0.000	0.005
N	747647	743078
Surveys	1	1

The sample includes household members eligible for the individual male and female surveys, aged 15–49. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.10: Diabetes, India 2015–16

	(1)	(2)
City	0.015***	0.012***
	(0.001)	(0.001)
Town	0.005***	0.006***
	(0.001)	(0.001)
Respondent's age/10		0.026**
		(0.009)
Respondent's age/10, squared		-0.011**
		(0.003)
Respondent's age/10, cubed		0.002***
		(0.000)
Female respondent		0.001
		(0.001)
Education of respondent: Primary		0.009***
		(0.001)
Education of respondent: Secondary		0.012***
		(0.001)
Education of respondent: Higher		0.009***
		(0.001)
ln(age of hh head)		0.001
		(0.001)
Female household head		0.002*
		(0.001)
Education of hh head: Primary		0.001
		(0.001)
Education of hh head: Secondary		0.003***
		(0.001)
Education of hh head: Higher		0.006***
		(0.001)
Country	India	India
Outcome mean	0.016	0.016
R^2	0.002	0.019
N	710204	705916
Surveys	1	1

The sample includes household members eligible for the individual male and female surveys, aged 15–49 and not pregnant. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.11: 8 years of schooling among 16-year-olds

	(1)	(2)	(3)	(4)
City	0.335***	0.231***	0.015	-0.042***
	(0.013)	(0.012)	(0.009)	(0.008)
Town	0.201***	0.145***	-0.067***	-0.076***
	(0.019)	(0.016)	(0.009)	(0.008)
Female child		-0.007		-0.007
		(0.008)		(0.005)
Female household head		0.086***		0.078***
		(0.009)		(0.008)
ln(age hh head)		0.151***		0.138***
		(0.012)		(0.012)
Education of hh head: Primary		0.114***		0.153***
		(0.010)		(0.008)
Education of hh head: Secondary		0.315***		0.274***
		(0.013)		(0.006)
Education of hh head: Higher		0.420***		0.364***
		(0.017)		(0.008)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.363	0.363	0.765	0.765
R^2	0.177	0.247	0.043	0.127
N	39133	38740	55333	55108
Surveys	29	29	3	3

The dependent variable is a dummy for whether the 16 year-old household member received at least 8 years of schooling. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.12: Number of births in past three years

	(1)	(2)	(3)	(4)
City	-0.195***	-0.105***	-0.044***	-0.014***
	(0.007)	(0.006)	(0.003)	(0.003)
Town	-0.110***	-0.055***	0.017***	0.027***
	(0.011)	(0.009)	(0.003)	(0.003)
Respondent's age/10		3.967***		5.377***
		(0.052)		(0.030)
Respondent's age/10, squared		-1.100***		-1.722***
		(0.017)		(0.010)
Respondent's age/10, cubed		0.092***		0.170***
		(0.002)		(0.001)
Education of respondent: Primary		-0.031***		-0.024***
		(0.005)		(0.003)
Education of respondent: Secondary		-0.136***		-0.044***
		(0.006)		(0.003)
Education of respondent: Higher		-0.257***		-0.149***
		(0.009)		(0.004)
Female household head		-0.150***		-0.025***
		(0.004)		(0.002)
ln(age of hh head)		-0.258***		-0.095***
		(0.006)		(0.004)
Education of hh head: Primary		-0.015**		-0.024***
		(0.005)		(0.003)
Education of hh head: Secondary		-0.038***		-0.039***
		(0.006)		(0.002)
Education of hh head: Higher		-0.074***		-0.015***
		(0.008)		(0.004)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.481	0.481	0.225	0.224
R^2	0.029	0.227	0.003	0.179
N	417251	412611	684800	680678
Surveys	29	29	3	3

All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.13: Current use of modern contraception

	(1)	(2)	(3)	(4)
City	0.151***	0.068***	-0.048***	-0.059***
	(0.008)	(0.008)	(0.013)	(0.012)
Town	0.094***	0.053***	-0.119***	-0.120***
	(0.014)	(0.012)	(0.013)	(0.013)
Respondent's age/10		0.071		1.511**
		(0.371)		(0.585)
Respondent's age/10, squared		0.000		-0.301
		(0.129)		(0.201)
Respondent's age/10, cubed		-0.003		0.018
		(0.015)		(0.023)
Education of respondent: Primary		0.102***		0.073***
		(0.006)		(0.010)
Education of respondent: Secondary		0.181***		0.075***
		(0.008)		(0.009)
Education of respondent: Higher		0.223***		0.010
		(0.015)		(0.014)
Female household head		0.027***		-0.061***
		(0.007)		(0.012)
ln(age of hh head)		0.020*		-0.078***
		(0.009)		(0.012)
Education of hh head: Primary		0.050***		0.022*
		(0.007)		(0.009)
Education of hh head: Secondary		0.047***		-0.010
		(0.008)		(0.009)
Education of hh head: Higher		0.059***		-0.040**
		(0.012)		(0.015)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.331	0.331	0.566	0.566
R^2	0.182	0.209	0.018	0.073
N	123944	122577	51000	50729
Surveys	28	28	3	3

The dependent variable is a dummy for whether the woman (aged 15–49) is currently using one of the following forms of contraception: pill, IUD, injections, diaphragm, condom, female sterilisation, male sterilisation, implants, female condom, foam and jelly, emergency contraception, other modern. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.14: Women who think beatings are justified

	(1)	(2)	(3)	(4)
City	-0.160***	-0.094***	-0.036**	-0.003
	(0.008)	(0.008)	(0.013)	(0.013)
Town	-0.071***	-0.042***	0.049***	0.059***
	(0.012)	(0.012)	(0.012)	(0.012)
Respondent's age/10		0.111*		0.127
		(0.047)		(0.076)
Respondent's age/10, squared		-0.047**		-0.037
		(0.016)		(0.025)
Respondent's age/10, cubed		0.005**		0.004
		(0.002)		(0.003)
Education of respondent: Primary		-0.023***		-0.016*
		(0.005)		(0.007)
Education of respondent: Secondary		-0.093***		-0.040***
		(0.007)		(0.006)
Education of respondent: Higher		-0.234***		-0.110***
		(0.009)		(0.009)
Female household head		-0.020***		-0.003
		(0.004)		(0.007)
ln(age of hh head)		-0.010*		-0.077***
		(0.005)		(0.009)
Education of hh head: Primary		0.001		-0.023***
		(0.005)		(0.007)
Education of hh head: Secondary		-0.031***		-0.055***
		(0.006)		(0.006)
Education of hh head: Higher		-0.072***		-0.103***
		(0.008)		(0.010)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.485	0.485	0.429	0.429
R^2	0.144	0.162	0.016	0.031
N	397168	392638	143967	143261
Surveys	29	29	3	3

The dependent variable is a dummy for whether it is the woman's (aged 15–49) opinion that a husband is justified in hitting or beating his wife when: she goes out without telling him, she argues with him, she refuses to have sex with him, or she burns the food. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.15: Ever experienced violence

	(1)	(2)	(3)	(4)
City	-0.023**	-0.020*	-0.043***	0.009
	(0.008)	(0.009)	(0.013)	(0.013)
Town	-0.030*	-0.031*	0.049***	0.062***
	(0.012)	(0.012)	(0.011)	(0.011)
Respondent's age/10		0.714***		0.980***
		(0.108)		(0.154)
Respondent's age/10, squared		-0.204***		-0.270***
		(0.034)		(0.048)
Respondent's age/10, cubed		0.019***		0.024***
		(0.004)		(0.005)
Education of respondent: Primary		0.062***		-0.018
		(0.008)		(0.011)
Education of respondent: Secondary		0.037***		-0.109***
		(0.009)		(0.008)
Education of respondent: Higher		-0.040**		-0.207***
		(0.015)		(0.012)
Female household head		-0.013		-0.007
		(0.007)		(0.009)
ln(age of hh head)		-0.031**		-0.116***
		(0.011)		(0.011)
Education of hh head: Primary		0.039***		-0.036***
		(0.008)		(0.010)
Education of hh head: Secondary		0.016		-0.069***
		(0.009)		(0.009)
Education of hh head: Higher		-0.047***		-0.118***
		(0.014)		(0.013)
Region	SSA	SSA	S Asia	S Asia
Outcome mean	0.296	0.296	0.306	0.305
R^2	0.066	0.078	0.005	0.052
N	115517	114406	62204	61813
Surveys	22	22	2	2

The sample includes all currently married women aged 15–49. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the DHS-cluster level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.16: Physically attacked in the last year

	(1)	(2)
City	0.020***	0.020***
	(0.005)	(0.005)
Town	-0.003	-0.003
	(0.005)	(0.005)
age divided by 10		0.034*
		(0.017)
age divided by 10, squared		-0.009*
		(0.004)
age divided by 10, cubed		0.001*
		(0.000)
female		-0.007*
		(0.003)
Primary education		-0.002
		(0.004)
Secondary education		0.004
		(0.005)
Tertiary education		0.004
		(0.007)
ln(hh size)		0.025***
		(0.003)
Region	SSA	SSA
Controls	No	Yes
Outcome mean	0.094	0.094
R^2	0.047	0.049
N	47839	47484
Surveys		

The dependent variable is a dummy for whether the respondent or any family members have been physically attacked in the past year. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the Afrobarometer primary sampling unit level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.17: Experienced home robbery in the past year

	(1)	(2)
City	0.047***	0.042***
	(0.007)	(0.007)
Town	0.019*	0.018*
	(0.009)	(0.009)
age divided by 10		0.010
		(0.026)
age divided by 10, squared		-0.002
		(0.006)
age divided by 10, cubed		0.000
		(0.000)
female		-0.011*
		(0.004)
Primary education		0.015*
		(0.007)
Secondary education		0.036***
		(0.008)
Tertiary education		0.036***
		(0.010)
ln(hh size)		0.038***
		(0.004)
Region	SSA	SSA
Controls	No	Yes
Outcome mean	0.278	0.278
R^2	0.045	0.049
N	47858	47503
Surveys		

The dependent variable is a dummy for whether the respondent or any family members have had something stolen from their house in the past year. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the Afrobarometer primary sampling unit level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.18: Feared crime in own home in the past year

	(1)	(2)
City	0.067***	0.064***
	(0.008)	(0.008)
Town	0.019	0.019
	(0.010)	(0.010)
age divided by 10		0.000
		(0.028)
age divided by 10, squared		-0.000
		(0.006)
age divided by 10, cubed		-0.000
		(0.000)
female		0.022***
		(0.004)
Primary education		-0.001
		(0.009)
Secondary education		-0.001
		(0.009)
Tertiary education		0.016
		(0.010)
ln(hh size)		0.009*
		(0.004)
Region	SSA	SSA
Controls	No	Yes
Outcome mean	0.317	0.316
R^2	0.076	0.077
N	47779	47426
Surveys		

The dependent variable is a dummy for whether the respondent or any family members have feared crime in their own home in the past year. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the Afrobarometer primary sampling unit level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

Table C.19: Felt unsafe walking in their neighbourhood in the past year

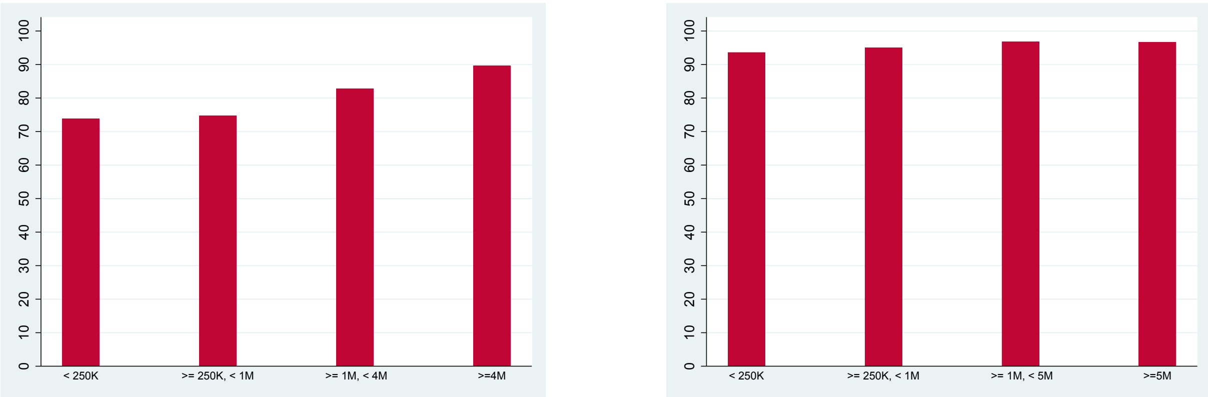
	(1)	(2)
City	0.089***	0.083***
	(0.009)	(0.009)
Town	0.018	0.017
	(0.011)	(0.011)
age divided by 10		0.005
		(0.029)
age divided by 10, squared		-0.002
		(0.006)
age divided by 10, cubed		0.000
		(0.000)
female		0.024***
		(0.004)
Primary education		0.016
		(0.008)
Secondary education		0.027**
		(0.009)
Tertiary education		0.046***
		(0.011)
ln(hh size)		0.019***
		(0.005)
Region	SSA	SSA
Controls	No	Yes
Outcome mean	0.373	0.373
R^2	0.077	0.079
N	47791	47439
Surveys		

The dependent variable is a dummy for whether the respondent or any family members have felt unsafe walking in their neighbourhood in the past year. All specifications follow equation (3) and use survey fixed effects. Standard errors are clustered at the Afrobarometer primary sampling unit level. P-values are indicated by * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

D City size class differentials

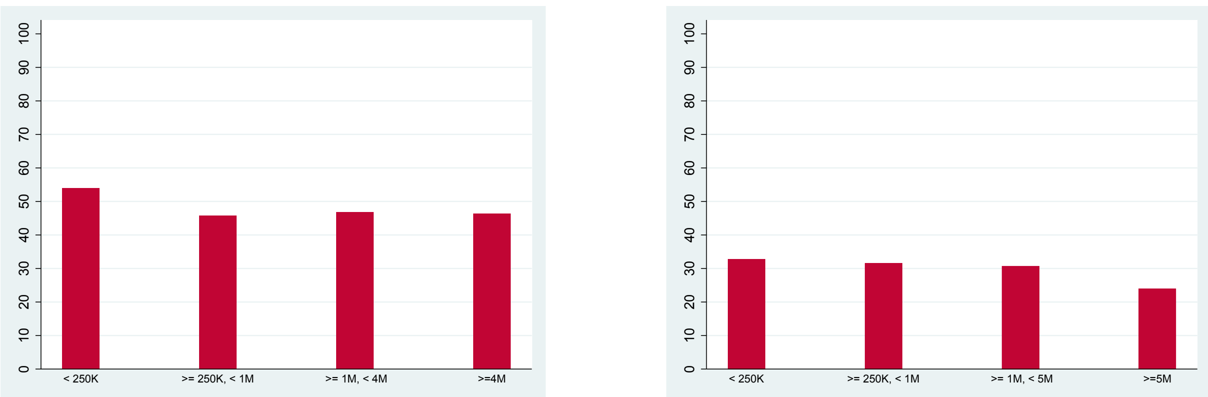
In this Appendix we provide a set of figures showing how a selected set of outcomes vary by city size class. Sometimes there is monotonic improvement as one moves up city size classes but many times there is not. Generally all differences are modest, except to some degree attitudes towards and experience with household violence outcomes. The key distinction seems to be cities overall versus rural areas and towns.

Figure D.1: Percentage of households with electricity, by city size class



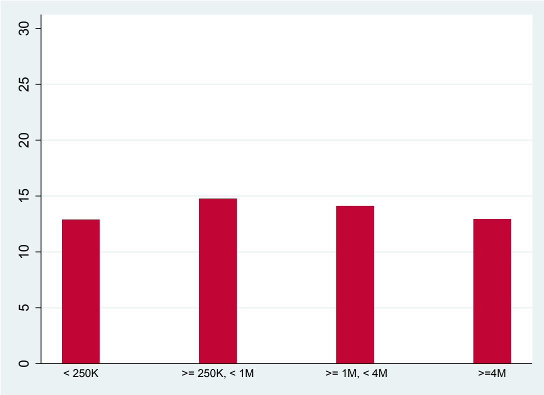
(a) SSA (b) S Asia

Figure D.2: Infant mortality rate, by city size class

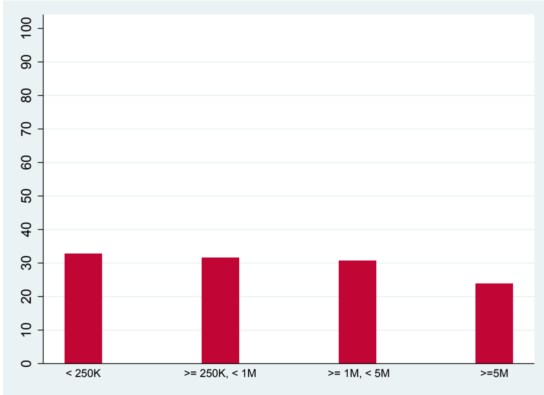


(a) SSA (b) S Asia

Figure D.3: Percentage of children with diarrhoea in the last 2 weeks, by city size class

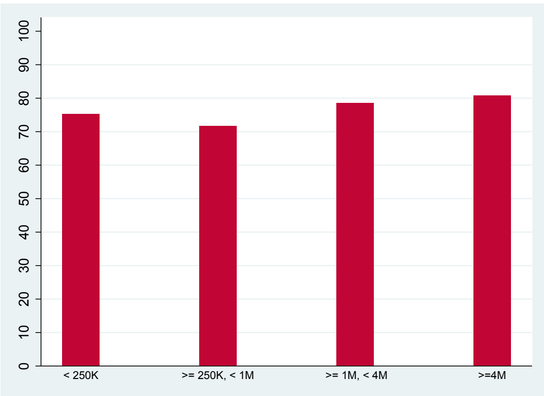


(a) SSA

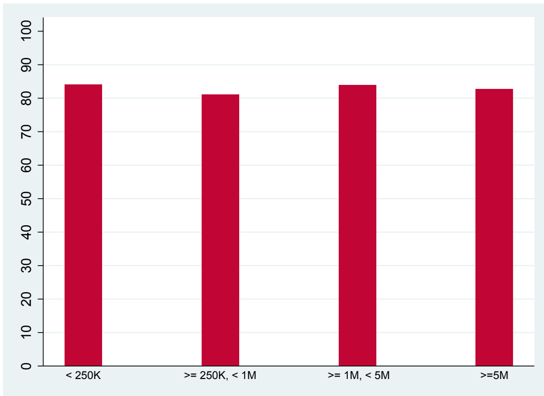


(b) S Asia

Figure D.4: Percentage of children with third DPT vaccination

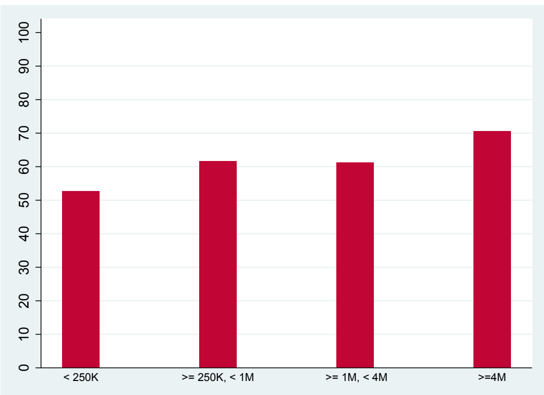


(a) SSA

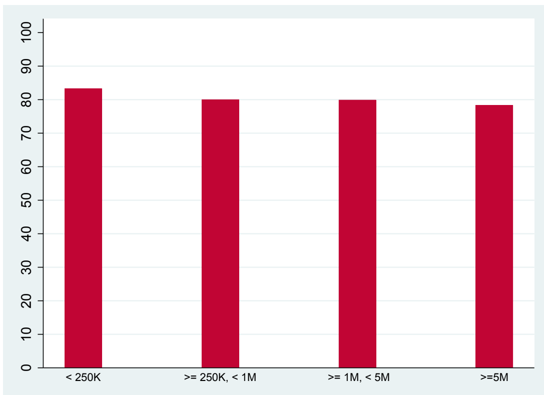


(b) S Asia

Figure D.5: Percentage of 16-year-olds with at least 8 years of education, by city size class

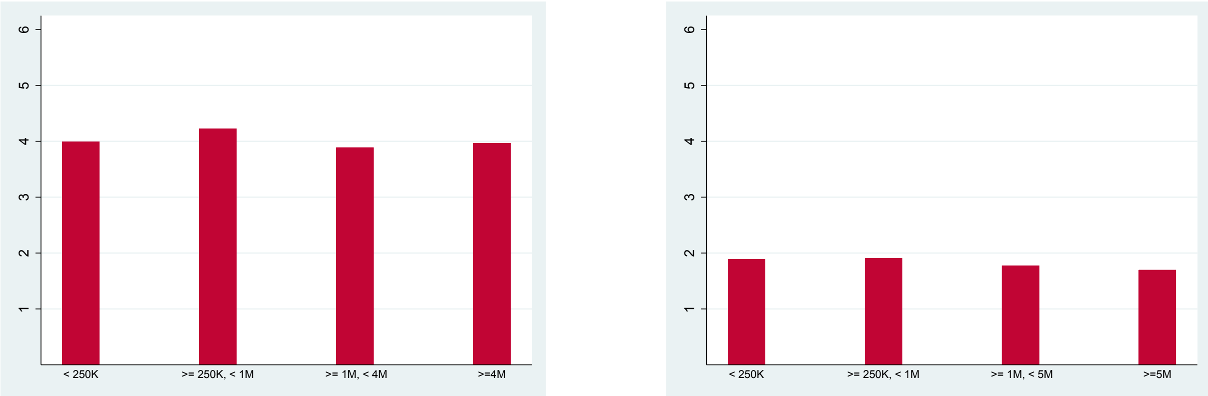


(a) SSA



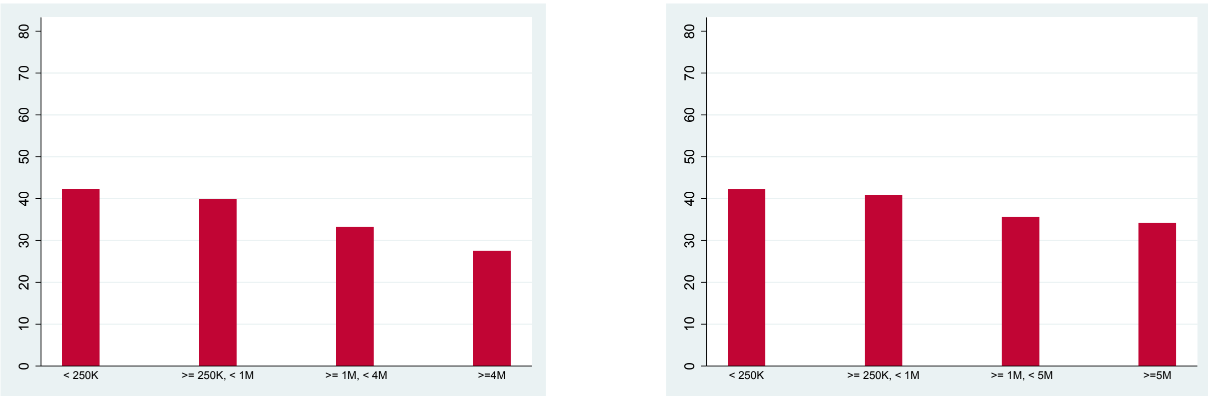
(b) S Asia

Figure D.6: Total fertility rate, per woman, by city size class



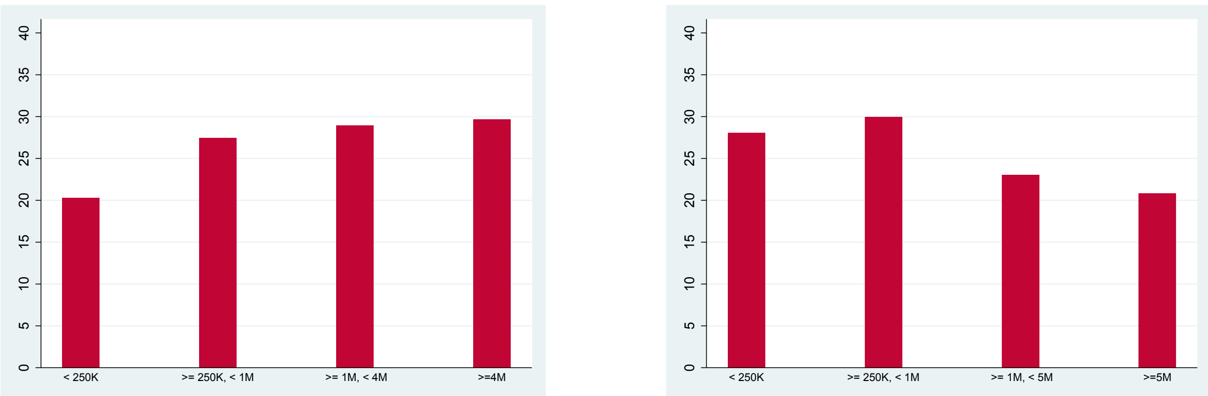
(a) SSA (b) S Asia

Figure D.7: Percentage of female respondents who believed wife beating was justified for at least one reason, by city size class



(a) SSA (b) S Asia

Figure D.8: Percentage of currently married female respondents who experienced physical spousal violence, by city size class



(a) SSA (b) S Asia

E Displacement correction

This section discusses the method to provide unbiased and consistent estimation on the differential in outcomes from living in a city or town relative to a rural area using displaced DHS cluster locations. This method is developed based on Karra and Canning (2018) where the authors presented a numerical integration method to address the estimation bias induced by displacement error in an explanatory variable.

Assume the true regression model is

$$Y_j = \alpha + \beta g(x_j) + \epsilon_j \tag{5}$$

where j indexes DHS clusters, Y_j indicates outcome, x_j is a tuple of (x_1, x_2) indicating the true location of a DHS cluster, $g(x_j)$ is an indicator function which equals to 1 if x_j falls into a zone z ($z \in \{city, town, rural\}$), ϵ_j is the error term.

In the context of DHS displacement, $g(x_j)$ is observed with errors. Denote m_j as the perturbed location (m_1, m_2) for the actual location x_j. Following the method proposed by Karra and Canning (2018), replacing $g(m_j)$ with $E(g(x_j)|m_j)$ gives unbiased and consistent estimation of $g(x_j)$. Specifically[20],

$$E(g(x)|m) = \int_X g(x)p(x|m)dx \tag{6}$$

$$= \int_X g(x)\frac{p(m|x)p(x)}{\int_X p(m|x)p(x)dx}dx \tag{7}$$

$E(g(x)|m)$ is the expected value of $g(x)$ given the observed displaced location m, given the knowledge of the displacement probability $p(m|x)$ from the true location x to the displaced location m, and the underlying probability that x is the true location, which is $p(x)$. As $g(\cdot)$ is an

[20] Subscript j is suppressed at the moment for simplicity.

indicator function, $E(g(x)|m)$ gives the probability of x falls in zone z. Calculating $E(g(x)|m)$ involves enumerating all possible true location x and working out $p(m|x)$ and $p(x)$. It is obviously impossible to enumerate every possible true location x as there are infinite number of x. Karra and Canning (2018) proposed to approximate $E(g(x)|m)$ using evenly spaced grids. Rather than using grids, we developed a ring based integration approach to approximate $E(g(x)|m)$. Given the potential area of true location x under DHS displacement procedure is a circle, ring based numeric integration approach is more customized to this displacement mechanics. It is more computationally efficient to implement in practice and has interesting graphic implications.

Our method involves drawing a buffer within upper bound of the displacement distance around a DHS cluster and divide it into a number of rings. To determine the probability that a cluster falls in zone z, we have

$$P_{zj} = E(g(x_j)|m_j) \approx \frac{\sum_s A_{zjs}/r_s}{\sum_s B_{js}/r_s} \tag{8}$$

where s indicates ring, r_s is the central radius of ring s, A_{zjs} is the population in the intersection of zone z with ring s, and B_{js} is the population in ring s. If the population is constant, then A_{zjs} and B_{js} are just the area of the intersection and the ring respectively.

Proof: This result follows immediately after converting the axis to a Polar coordinate system. In this coordinate system, the perturbed location is the origin, $x_1 = r\cos\theta, x_2 = r\sin\theta$, where r is the distance between true location x and perturbed location m. θ is the angle of a line that connects the true location (x_1, x_2) and the perturbed location (m_1, m_2).

We obtained DHS survey data with and without displacement for 5 countries, including Burkina Faso 1998-1999, Ghana 1998, Mali 2001, Niger 1998 and Togo 1998. The five surveys give a sample of 39796 respondents. This allows comparing the estimated coefficients using the true location, displaced location and the expected locations. We demonstrate the effectiveness of this method using the simple linear regression in equation (5). We focus on one outcome variable from the female survey: whether the respondent has received higher education . The coefficient of interest is β, which is the differentials from living in a city relative to a town/rural area [21].

To evaluate the effectiveness of correction method, we regress Y_i on g(x) (true location), g(m) (displaced location) and E(g(x|m)) (expected location) respectively. To calculate $E(g(x_j)|m_j)$, We create 10 concentric rings with intervals of 200 meters around each perturbed point m, and then calculate A_{zjs} and B_{js}, and finally apply the formula in equation (8). In order to obtain standard errors, we adopt a sub-sampling procedure as in Politis et al. (1999). To be specific, we first randomly choose 10% of the full sample without replacement to run above three regressions. We then repeat the process 1000 times, which gives three arrays of $\hat{\beta}^{\text{true}}$, $\hat{\beta}^{\text{displaced}}$, and $\hat{\beta}^{\text{expected}}$.

A comparison of the distribution of the three coefficients is shown in Figure E.1. The black line shows the distribution of $\hat{\beta}^{\text{true}}$ estimated with the true location. The red line shows the $\hat{\beta}^{\text{displaced}}$ when using the displaced locations. The short vertical lines near the x-axis show the mean of the estimates. The displacement yields an upward bias. Applying the correction procedure leads us to the blue line, where the mean of the estimates $\hat{\beta}^{\text{expected}}$ is very close to the mean of $\hat{\beta}^{\text{true}}$. This illustrates that the correction procedure has contributed to eliminating the displacement bias in the estimation.

Note that the displacement is particularly severe for the DHS clusters close to the boundary between cities and town/rural area, as these clusters are more likely to be displaced into wrong zones. To test our correction procedure in this more stringent case, we restrict the sample to DHS clusters within 5km of the boundary between cities and town/rural area. Figure E.2 shows the result. Similar to the result in the main sample, the estimation based on the expected locations has improved the estimation towards the estimates based on the true locations. Of course, this correction procedure does not come free given the noise induced by the DHS displacement. The cost in applying the correction is an enlarged confidence interval, reflected by the fatter tails of the distribution of estimates using expected locations.

[21] We can easily expand this to a multivariate framework. Adding a town dummy to the regression allows estimating city-rural and town-rural differential separately.

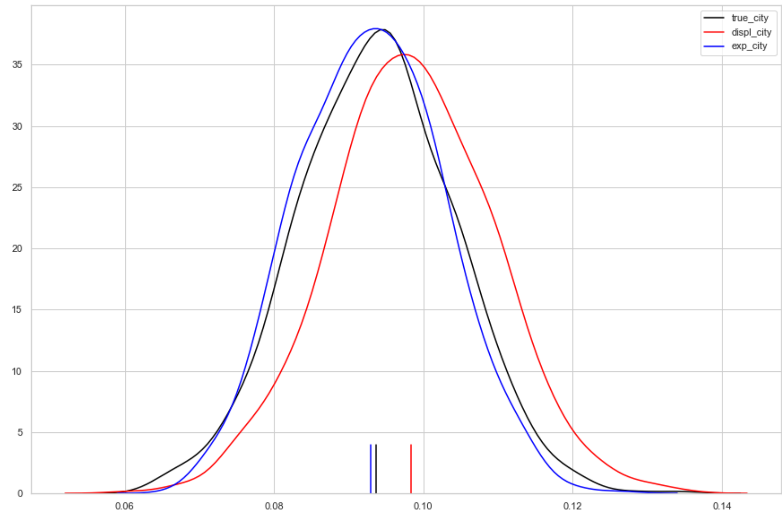


Figure E.1: Evaluate the correction procedure: full sample

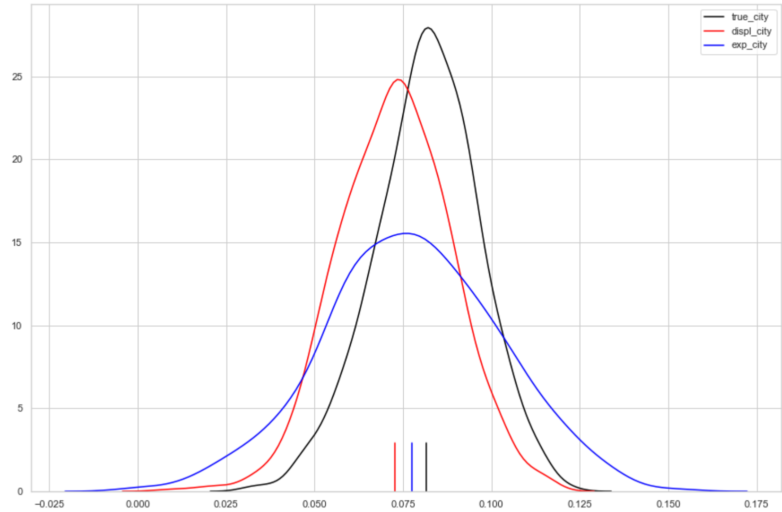


Figure E.2: Evaluate the correction procedure: within 5km of city boundaries

References

Ades, Alberto F and Edward L Glaeser (1995) "Trade and circuses: explaining urban giants," The Quarterly Journal of Economics, Vol. 110, No. 1, pp. 195–227.

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani (2019) "The economic effects of density: A synthesis," Journal of Urban Economics, Vol. 111, pp. 93–107.

Altonji, Joseph G and Richard K Mansfield (2018) "Estimating group effects using averages of observables to control for sorting on unobservables: School and neighborhood effects," American Economic Review, Vol. 108, No. 10, pp. 2902–46.

Bacolod, Marigee, Bernardo S Blum, and William C Strange (2009) "Skills in the city," Journal of Urban Economics, Vol. 65, No. 2, pp. 136–153.

Balk, Deborah, Mark Montgomery, Hasim Engin, Natalie Lin, Elizabeth Major, and Bryan Jones (2019) "Urbanization in India: Population and Urban Classification Grids for 2011," Data, Vol. 4, No. 1, p. 35, URL: http://dx.doi.org/10.3390/data4010035, DOI: 10.3390/data4010035.

Balk, Deborah, Thomas Pullum, Adam Storeygard, Fern Greenwell, and Melissa Neuman (2004) "A spatial analysis of childhood mortality in West Africa," Population, Space and Place, Vol. 10, No. 3, pp. 175–216.

Balk, D.L., U. Deichmann, G. Yetman, F. Pozzi, S.I. Hay, and A. Nelson (2006) Determining Global Population Distribution: Methods, Applications and Data, pp. 119–156: Elsevier, URL: http://dx.doi.org/10.1016/S0065-308X(05)62004-0, DOI: 10.1016/s0065-308x(05)62004-0.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang (2017) "Roads, Railroads, and Decentralization of Chinese Cities," The Review of Economics and Statistics, Vol. 99, No. 3, pp. 435–448, DOI: 10.1162/REST_a_00660.

Behrens, Kristian and Frederic Robert-Nicoud (2015) "Agglomeration theory with heterogeneous agents," in Handbook of regional and urban economics, Vol. 5: Elsevier, pp. 171–245.

Bongaarts, John and John Casterline (2013) "Fertility Transition: Is sub-Saharan Africa Different?" Population and development review, Vol. 38, No. Suppl 1.

Brueckner, Jan K and Somik V Lall (2015) "Cities in developing countries: fueled by rural–urban migration, lacking in tenure security, and short of affordable housing," in Handbook of regional and urban economics, Vol. 5: Elsevier, pp. 1399–1455.

Bryan, Gharad and Melanie Morten (2019) "The aggregate productivity effects of internal migration: Evidence from indonesia," Journal of Political Economy, Vol. 127, No. 5, pp. 000–000.

Center For International Earth Science Information Network-CIESIN-Columbia University (2017) "Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 10," Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), URL: https://doi.org/10.7927/H4PG1PPM, DOI: 10.7927/H4PG1PPM.

Center For International Earth Science Information Network-CIESIN-Columbia University; International Food Policy Research Institute-IFPRI; The World Bank; Centro Internacional De Agricultura Tropical-CIAT (2011) "Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Density Grid," Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), URL: http://sedac.ciesin.columbia.edu/data/set/grump-v1-population-density, DOI: 10.7927/h4r20z93.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio (2017) "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States," Journal of Urban Economics, Vol. 98, pp. 17–49.

Cleland, John, George Bicego, and Greg Fegan (1992) "Socioeconomic inequalities in childhood mortality: the 1970s to the 1980s," Health Transition Review, Vol. 2, No. 1, pp. 1–18, URL: http://www.jstor.org/stable/40652030.

Combes, Pierre-Philippe, Sylvie Démurger, Shi Li, and Jianguo Wang (2019) "Unequal migration and urbanisation gains in China," Journal of Development Economics.

Corbane, Christina, Aneta Florczyk, Martino Pesaresi, Panagiotis Politis, and Vasileios Syrris (2018) "GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014), R2018A," European Commission, Joint Research Centre (JRC), DOI: doi:10.2905/jrc-ghsl-10007.

Corbane, Christina, Martino Pesaresi, Thomas Kemper, Panagiotis Politis, Aneta J. Florczyk, Vasileios Syrris, Michele Melchiorri, Filip Sabo, and Pierre Soille (2019) "Automated global delineation of human settlements from 40 years of Landsat satellite data archives," Big Earth Data, Vol. 3, pp. 140–169, DOI: 10.1080/20964471.2019.1625528.

Corker, Jamaica (2017) "Fertility and Child Mortality in Urban West Africa: Leveraging Geo-Referenced Data to Move Beyond the Urban/Rural Dichotomy," Population, Space and Place, Vol. 23, DOI: 10.1002/psp.2009.

Croft, Trevor N., Aileen M. J. Marshall, Courtney K. Allen et al. (2018) Guide to DHS Statistics: ICF.

Davis, James C and J Vernon Henderson (2003) "Evidence on the political economy of the urbanization process," Journal of urban economics, Vol. 53, No. 1, pp. 98–125.

Diamond, Rebecca (2016) "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," American Economic Review, Vol. 106, No. 3, pp. 479–524.

Dorélien, Audrey, Deborah Balk, and Megan Todd (2013) "What Is Urban? Comparing a Satellite View with the Demographic and Health Surveys," Population and Development Review, Vol. 39, No. 3, pp. 413–439, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1728-4457.2013.00610.x, DOI: 10.1111/j.1728-4457.2013.00610.x.

Ferré, Céline, Francisco H.G. Ferreira, and Peter Lanjouw (2012) "Is There a Metropolitan Bias? The relationship between poverty and city size in a selection of developing countries," The

World Bank Economic Review, Vol. 26, No. 3, pp. 351–382, URL: http://dx.doi.org/10.1093/wber/lhs007, DOI: 10.1093/wber/lhs007.

Florczyk, Aneta, Christina Corbane, Daniele Ehrlich, Sergio Freire, Thomas Kemper, Luca Maffenini, Michele Melchiorri, Martino Pesaresi, Panagiotis Politis, Marcelo Schiavina, Filip Sabo, and Luigi Zanchetta (2019) "GHSL Data Package 2019,"Technical Report EUR 29788 EN, Publications Office of the European Union.

Freire, Sergio, Kytt MacManus, Martino Pesaresi, Erin Doxsey-Whitfield, and Jane Mills (2016) "Development of new open and free multi-temporal global population grids at 250 m resolution," in Geospatial Data in a Changing World, Association of Geographic Information Laboratories in Europe (AGILE).

Gollin, Douglas, Martina Kirchberger, and David Lagakos (2018) "In search of a spatial equilibrium in the developing world,"Technical report, Trinity College Dublin.

Gollin, Douglas, David Lagakos, and Michael E Waugh (2013) "The agricultural productivity gap," The Quarterly Journal of Economics, Vol. 129, No. 2, pp. 939–993.

Henderson, J Vernon and Sebastian Kriticos (2018) "The development of the african system of cities," Annual Review of Economics, Vol. 10, pp. 287–314.

Henderson, J Vernon, Dzhamilya Nigmatulina, and Sebastian Kriticos (2019) "Measuring urban economic density."

Jedwab, Remi and Adam Storeygard (2017) "The Average and Heterogeneous Effects of Transportation Investments: Evidence from sub-Saharan Africa 1960-2010," Discussion Papers Series, Department of Economics, Tufts University 0822, Department of Economics, Tufts University.

Kahn, Matthew E and Randall Walsh (2015) "Cities and the Environment," in Handbook of regional and urban economics, Vol. 5: Elsevier, pp. 405–465.

Karra, Mahesh and David Canning (2018) "Consistent Estimation with Deliberate Measurement Error to Protect Subject Confidentiality: An Application to Perturbed Location Data," Working Paper 003, Boston University Global Development Policy Center Human Capital Initiative.

Lagakos, David and Michael E Waugh (2013) “Selection, agriculture, and cross-country productivity differences,” American Economic Review, Vol. 103, No. 2, pp. 948–80.

Lall, Somik V, Christopher Timmins, and Shouyue Yu (2009) Connecting lagging and leading regions: The role of labor mobility: The World Bank.

Ministry of Health, Community Development, Gender, Elderly, and Children (MoHCDGEC) and Ministry of Health (MoH/Zanzibar) and National Bureau of Statistics (NBS) and Office of Chief Government Statistician (OCGS/Zanzibar) and ICF (2016) Tanzania Demographic and Health Survey and Malaria Indicator Survey 2015-2016: MoHCDGEC, MoH, NBS, OCGS, and ICF.

Pesaresi, Martino, Aneta Florczyk, Marcello Schiavina, Michele Melchiorri, and Luca Maffenini (2019) “GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A, multitemporal (1975-1990-2000-2015), R2019A,” European Commission, Joint Research Centre (JRC), DOI: 10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf (1999) Subsampling, New York: Springer.

Potere, David and Annemarie Schneider (2007) “A critical look at representations of urban areas in global maps,” GeoJournal, Vol. 69, No. 1–2, pp. 55–80.

Potere, David, Annemarie Schneider, Shlomo Angel, and Daniel L. Civco (2009) “Mapping urban areas on a global scale: which of the eight maps now available is more accurate?” International Journal of Remote Sensing, Vol. 30, No. 24, pp. 6531–6558, URL: https://doi.org/10.1080/01431160903121134, DOI: 10.1080/01431160903121134.

Quintero, Luis E and Mark Roberts (2018) Explaining spatial variations in productivity: Evidence from Latin America and the Caribbean: The World Bank.

Schiavina, Marcello, Sergio Freire, and Kytt MacManus (2019) "GHS population grid multitemporal (1975-1990- 2000-2015), R2019A," European Commission, Joint Research Centre (JRC), DOI: doi:10.2905/0C6B9751- A71F-4062-830B-43C9F432370F.

Tombe, Trevor and Xiaodong Zhu (2019) "Trade, migration, and productivity: A quantitative analysis of china," American Economic Review, Vol. 109, No. 5, pp. 1843–72.

United Nations, Department of Economic and Social Affairs, Population Division (2018) "World Urbanization Prospects: The 2018 Revision," Online Edition.

Young, Alwyn (2013) "Inequality, the urban-rural gap, and migration," The Quarterly Journal of Economics, Vol. 128, No. 4, pp. 1727–1785.

Getting in touch with the EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: https://europa.eu/european-union/contact_en

Finding information about the EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: http://eur-lex.europa.eu

Open data from the EU

The EU Open Data Portal (http://data.europa.eu/euodp/en) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

KN-03-20-034-EN-N

Publications Office
of the European Union

ISBN: 978-92-76-15239-2