Elsevier Editorial System(tm) for The Lancet
Manuscript Draft

Manuscript Number: THELANCET-D-18-07702R2

Title: Gender norms and health: insights from global survey data

Article Type: Invited Series

Keywords: Gender norms
Global health
Health inequities
Gender equality
Social determinants of health
Proxy measures
Methods

Corresponding Author: Dr. Ann Weber, PhD

Corresponding Author's Institution: Stanford University School of Medicine

First Author: Ann Weber, PhD

Order of Authors: Ann Weber, PhD; Beniamino Cislaghi, PhD; Valerie Meausoone, BS; Safa Abdalla, MD; Iván Mejía-Guevara, PhD; Pooja Loftus, MS; Emma Hallgren, PhD; Ilana J Seff, MPH; Lindsay Stark, DrPh; Cesar G Victora, MD; Romina Buffarini, PhD; Aluísio J Barros, MD, PhD; Ben Domingue, PHD; Devika Bhushan, MD; Ribhav Gupta, BS; Jason Nagata, MD; Holly B Shakya, PhD; Linda Richter, PhD; Shane A Norris, PhD; Thoai Ngo, PhD; Sophia Chae, PhD; Nicole Haberland, MPH; Katherine McCarthy, MPH; Mark Cullen, MD; Gary L Darmstadt, MD

Abstract: Despite global commitments to achieving gender equality and improving health and well-being for all, quantitative data and methods to precisely estimate the effect of gender norms on health inequities are under-developed. Nonetheless, existing global, national, and sub-national data provide key opportunities for testing associations between gender norms and health. Using innovative approaches to analysing proxies for gender norms, we generated evidence that gender norms impact the health of women and men across life stages, health sectors, and world regions. Six case studies demonstrated that: 1) gender norms are complex and may intersect with other social factors to impact health over the life course; 2) early gender-normative influences by parents and peers may have multiple and differing health consequences for girls and boys; 3) non-conformity with, and transgression of, gender norms may be harmful to health, in particular when they trigger negative sanctions; and 4) the impact of gender norms on health can be context-specific, demanding care when designing effective gender-transformative health policies and programs. Limitations of survey-based data are described that resulted in missed opportunities for exploring certain populations and domains. Recommendations for optimising and advancing research on the health impacts of gender norms are made.

Paper 2 of *The Lancet* Series on Gender Equality, Norms and Health

Gender norms and health: insights from global survey data

 # Words: Abstract: 196; body: 5,566

# of Objects: 5

Authors:

Ann Weber, PhD[1]; Beniamino Cislaghi, PhD[2]; Valerie Meausoone, BS[3]; Safa Abdalla, MD[1]; Iván Mejía-Guevara, PhD[3]; Pooja Loftus, MS[3]; Emma Hallgren, PhD [3]; Ilana Seff, MPH[4]; Lindsay Stark, DrPh[5]; Cesar Gomes Victora, MD, PhD[6]; Romina Buffarini, PhD[6]; Aluísio J. D. Barros, MD, PhD[6]; Ben Domingue, PhD[7]; Devika Bhushan, MD[1]; Ribhav Gupta, PhD[1]; Jason M. Nagata, MD[8;] Holly B. Shakya, PhD[9]; Linda Richter, PhD[10]; Shane A Norris, PhD[10]; Thoai Ngo, PhD[11]; Sophia Chae, PhD[11]; Nicole Haberland, MPH[11]; Katharine McCarthy, MPH[11]; Mark Cullen, MD[3]; Gary L. Darmstadt, MD[1,3]; for the Gender Equality, Norms and Health Steering Committee*

1 Stanford University
Department of Pediatrics, Stanford University School of Medicine
1265 Welch Road, Stanford, CA 94305 USA

2 London School of Hygiene & Tropical Medicine
15-17 Tavistock Place
London WC1H 9SH
United Kingdom

3 Center for Population Health Sciences
Stanford University School of Medicine
1070 Arastradero Rd, Palo Alto, CA 94304 USA

4 Columbia University
Department of Population and Family Health
Mailman School of Public Health
722 West 168th St. NY, NY 10032 USA

5 Brown School at Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130

6 Federal University of Pelotas, Brazil
International Center for Equity in Health
Rua Deodoro, 1120
96020-220 Pelotas, Rs, Brazil

7 Stanford University
Graduate School of Education
485 Lasuen Mall, Stanford, CA 94305 USA

8 University of California San Francisco

Department of Pediatrics
3333 California Street
San Francisco CA 94143

9 University of California San Diego
Department of Medicine, Center on Gender Equity and Health (GEH)
9500 Gilman Drive
La Jolla, CA 92093 USA

10 University of Witwatersrand
Wits School of Public Health
Private Bag 3
Wits 2050
Johannesburg, South Africa

11 Population Council
One Dag Hammarskjold Plaza
New York, NY 10017 USA

* Members listed at the end of the manuscript

Corresponding Author: Ann Weber, PhD[1]

Email address: annweber@stanford.edu

Funding

This Series was funded by the Gender Equality, Integrated Delivery, HIV, Nutrition, Family Planning and Water Sanitation and Hygiene Program Strategy Teams at the Bill & Melinda Gates Foundation, and the United Arab Emirates Ministry of Foreign Affairs and International Cooperation, through grants to Stanford University. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data and had final responsibility to submit the paper for publication.

Disclaimer

The views and opinions expressed in this paper are those of the authors and do not reflect the official position of any of the organizations for which the authors work.

Abstract

Despite global commitments to achieving gender equality and improving health and well-being for all, quantitative data and methods to precisely estimate the effect of gender norms on health inequities are under-developed. Nonetheless, existing global, national, and sub-national data provide key opportunities for testing associations between gender norms and health. Using innovative approaches to analysing proxies for gender norms, we generated evidence that gender norms impact the health of women and men across life stages, health sectors, and world regions. Six case studies demonstrated that: 1) gender norms are complex and may intersect with other social factors to impact health over the life course; 2) early gender-normative influences by parents and peers may have multiple and differing health consequences for girls and boys; 3) non-conformity with, and transgression of, gender norms may be harmful to health, in particular when they trigger negative sanctions; and 4) the impact of gender norms on health can be context-specific, demanding care when designing effective gender-transformative health policies and programs. Limitations of survey-based data are described that resulted in missed opportunities for exploring certain populations and domains. Recommendations for optimising and advancing research on the health impacts of gender norms are made.

Key Messages

1. Existing survey-based data can be harnessed to generate new evidence of the pervasive influence of gender norms on the health and well-being of girls, boys, women, and men across a range of health-related outcomes and the life course in high, middle, and low-income countries. While these data may be inadequate for making causal claims of the impact of specific gender norms on health, the data were sufficient to expose important gendered pathways to health and well-being. Additional opportunities remain to build on this evidence and generate new hypotheses with survey-based data.
2. By applying diverse analytical methods to different types of proxy measures for gender norms, we demonstrated that:
    a. Gender norms are complex and may intersect with other social factors to impact health over the life course;
    b. Gender-normative influences by parents and peers start early, and may have multiple short- and long-term health consequences that differ for girls and boys;
    c. Non-conformity and transgression of gender norms can be harmful to health, in particular when they trigger negative sanctions; and
    d. Gender norms are often context-specific, demanding a deeper understanding to design effective gender-transformative policies and programmes.
3. Existing survey-based data can introduce or perpetuate bias when used for studying the impact of gender norms on health:
    a. Reliance on sex-disaggregated data can result in misclassification of gender and ignores trans-gender and non-binary experiences.
    b. Datasets include rich gender-related attitude data or health-related data, but rarely both;
    c. Data are limited or non-existent for who enforces norms, how they are enforced, or what sanctions transgressors of norms may face.

d. Global datasets are generally not powered to study how gender norms intersect with strata of other social determinants of health (e.g., wealth, religion, and ethnicity) and may be missing data for entire demographic groups (e.g., boys and men, children 6-14 years, women over 49 years, gender minorities) or world regions.

e. Questions are often unbalanced by sex of the respondent (e.g., only women are asked about child health and care) and phrasing of questions frequently revealed underlying gender biases in research.

4. Future development of quantitative proxy measures for gender norms would benefit from mixed methods that utilise qualitative research to unpack the origins, preservation, and shifts in gender norms and their links with health outcomes.

5. Going forward, data on all facets of gender, including data for gender minorities, are necessary in future surveys with the above limitations addressed. To achieve these goals, collaborations are needed at multiple levels:

    a. Across disciplines to provide a conceptual bridge for effective use of data that aligns around an evidence-based research agenda;

    b. Between domain experts and gender scholars, survey designers and analysts, and community partners and policy makers to generate data systems that will enable studying health at the intersection of gender and other social determinants; and

    c. Across global data collection organisations to set standards for measuring gender, gender norms, and key demographic characteristics.

Introduction

Gender equality is a foundational human right, reflected in Sustainable Development Goal (SDG) 5, and a necessary means to achieve other SDGs, including 3, to "ensure healthy lives and promote well-being for all."[1,2] Mixed-methods studies document the consequences of gender inequality for women's and men's health.[3–6] However, quantitative data and methods are under-developed to precisely estimate these consequences and study how gender norms may contribute to health inequities. Nonetheless, existing survey-based data can be leveraged to gain important insights into pathways from gender norms to health.

Gender norms are society's spoken and unspoken rules about acceptable ways of being a girl or a boy, a woman or a man – how they should behave, look, and even think or feel. Gender norms are perpetuated and challenged in families, communities, schools, workplaces, institutions and the media.[3,5,7–9] These expectations start early and powerfully shape individuals' attitudes, opportunities, experiences, and behaviours, with important health consequences throughout the life course.[10]

Quantifying the effect of gender inequalities on health is challenging, partly because differences related to sex- (e.g., biological factors, including chromosomal, hormonal, and biomechanical) and gender (e.g., culturally-defined constructs associated with being female or male) are intertwined.[11–14] Globally, women outlive men by 2-4 years on average, but girls and women have a higher burden of some disabilities and morbidities.[2,15–18] These differences cannot be explained by sex alone, which we demonstrate with the 2016 Global Burden of Disease data,[19] extending work by Snow (2008).[20] We identified 15 causes of disability-adjusted life years (DALYs) that most disproportionately affected females (Figure 1a) or males (Figure 1b) globally. The >40:1 female-to-male DALY ratio from breast cancer is primarily sex-driven, whereas the ~3:1 female-to-male DALY ratio from eating disorders reflects gender-related factors.[3] Higher road traffic injuries among males, explaining nearly 4% of their all-cause age-standardised DALYs, also reflects male gender norms pertaining to driving, risk-taking, and alcohol use.[21] Sex/gender also intersect with other social factors to impact DALY ratios. For example, given differential exposures

within gendered occupations,[10] women are more vulnerable to Ebola (from nursing) in low Socio-Demographic Index (SDI) countries and men to pneumoconiosis (from mining) in high-SDI countries.[11,22]

From over a dozen case studies involving secondary analyses of existing global, national, and sub-national datasets, we selected six to present here (Table 1) based on conceptual and practical considerations (see Appendix 8 for the selection process). Conceptually, we aimed to study a range of gendered pathways to health for which evidence exists, as framed by Heise, Greene et al.[10] Our analyses were informed by feminist sociological theories of how gender norms contribute to shaping an unequal gender system that can be harmful to both women, men, boys and girls. [13,23–25] We sought to include pathways across the life course, around the world, and for diverse mental and physical health-related outcomes, despite challenges in data quality and operationalising gender norms. Following the case studies, we reflect on data opportunities and limitations, concluding with recommendations for optimising research on health impacts of gender norms.

Gendered pathways to health

We rely on sex-disaggregated data, recognising that sex and gender typically are conflated in surveys.[26,27] Additionally, existing survey data do not systematically measure gender norms, so we created proxies by aggregating individual-level data to the level of influential social or reference groups (e.g. peers). With the exception of studies 2 and 3, we aggregated gendered behaviours (what women/girls and men/boys do) or attitudes (what people believe women or men should do) to the level of a community, community cluster, or school. We then tested different pathways between gender norms and health. When data allowed, we tested how gender interacted with other analytical categories (e.g. wealth or religion) in shaping health-related social disadvantages. In case studies 1 and 5, we contrasted aggregated behaviours or attitudes for males and females to ask: "what can these differences tell us about gender norms and their implications for health?" In case studies 5 and 6, we asked of between-group variation: "can we detect differences in individual health by the strength of the gender-normative environment?" In case studies 4

and 6, we contrasted individual behaviour with that of groups to ask: "can non-conformity with, or transgression of, the norm impact individual health—for example, can it result in harm?" Finally, in case 5, we contrasted group-level attitudes (what people should do) with the corresponding behaviour (what people actually do) to ask: "can the discordance between them impact individual health?" Only in case studies 2 and 3 do we use individual-level data for the norm, taking advantage of the normative questions: "what do you think others think about you?" to explore gender differences and ask: "can a person's belief in what others think of them affect their health?"

For each case study presented below, we link the case to a gendered pathway, including key literature; describe the data, gender norm proxy measure, and analytic approach; and present key results and insights. The case studies are arranged by life stage, from childhood, to adolescence, to early adulthood.

Case study 1. Care-seeking for childhood illness in Ethiopia

Restrictive gender norms can affect young children's health. For example, when girls are seen as a lesser financial asset than boys, parents might invest less in girls' health and education,[28–31] reflected in differences in access to care for common childhood illnesses.[32] We used geospatial information available in the Demographic and Health Survey (DHS) for Ethiopia in 2011 to examine differences in care-seeking for girls and boys <5 years (n=3,161 children in 544 villages), which we hypothesised varied within country by geographic and sociodemographic contexts.[33,34] Care-seeking was defined as medical care sought from a certified medical practitioner for symptoms of pneumonia, fever, or diarrhoea (available disease indicators) in the previous two weeks.

We aggregated individual care-seeking behaviour using geospatial hierarchical cluster analysis [35] identifying spatially proximal clusters of communities with significantly higher (hot spots) and lower (cold spots) care-seeking than the national average, separately for girls, boys, and the differential (boys minus girls) (Appendix 1). We created a gender norms proxy of gender preference in care-seeking by

assigning a yes/no indicator to communities in hot spots for differential care-seeking. We tested whether key community-level characteristics (e.g., socio-economic status, dominant religion, and vaccination rates) predicted this proxy measure.

Hot and cold spots were mapped separately for girls and boys (Figure 2). Sex-specific maps were overlaid with spatial distributions of increasingly wealthy (panels 2a and 2b) and Muslim (panels 2c and 2d) households in communities (see Appendix 1 for factor selection). Clusters of hot (or cold) spots for girls and hot (or cold) spots for boys appear in the same geographic areas, except for a cluster of hot spots for boys in the east, for which there is no equivalent for girls and where communities appear wealthier and majority Muslim. In adjusted logistic regressions of sex-specific hot spots, we found that majority Muslim (>50% of households) communities were associated with increased odds of being care-seeking hot spots for boys but decreased odds for girls compared to communities with <50% Muslim households (Appendix Table A1.4). Differential care-seeking hot spots favouring boys had a very large and significant association with majority Muslim compared to minority Muslim communities (OR=18·2, 95% CI 8·72, 40·7; p-value<0·0001) (Appendix Table A1.4). Differential care-seeking favouring boys was also associated with mostly wealthy (>50% of households) communities, but the association was weaker and not statistically significant (OR=2·67, 95% CI 0·95, 7·46; p-value 0·062). We found no clear evidence for interaction between wealth and religion on care-seeking hot spots.

These findings suggest that, unlike reports from elsewhere,[36] poverty did not drive lower care-seeking for girls in Ethiopia. Our findings, however, are consistent with reports of son preference in other contexts.[37,38] Notably, preferential care-seeking for boys in Ethiopia was very strongly associated with Muslim majority communities. Evidence of care seeking in favour of boys in geographically focused communities, regardless of socioeconomic status, suggests that equal access to care is insufficient in achieving gender equality and highlights the importance of local contextual variation when addressing gender norms in programming and policy.

Case studies 2 and 3. Adolescent weight control and mental health in South Africa and Brazil

Gender norms learned in the family[7,39–41] are later reinforced or challenged in the community, at school, and by the media.[9,10] Evidence suggests that internalisation of gender norms and their influence on health-related behaviours might be especially powerful during adolescence,[7–9,41–43] when important biological and psychological changes occur and many health-related behaviours are adopted.[44,45] We examine pathways through which normative pressures from parents and peers may contribute to adolescents' gendered health behaviours and differential health outcomes. We present two complementary studies together as they offered unique data on individuals' perceptions of norms around body image.

Case 2:

Known manifestations of weight concerns—for example, eating disorders—are highly gendered globally, primarily affecting girls.[3,46,47] We used prospective cohort data from South Africa (Birth-to-20)[48] to examine how early normative pressures from peers affected adolescents' later weight control behaviour, and how this association differed by sex/gender and social context. The data are from mostly Black children (N=3,273) born in Soweto-Johannesburg in the early 1990s, during a period of rapid urbanisation[48] and simultaneous emergence of eating disorders among Black girls.[49]

The gender norms measure was adolescent boys' or girls' perceptions of peers' approval of their appearance (measured on a scale of 0-never to 4-always). Adjusted linear regression models used sex-disaggregated data from ages 13, 17, and 22 years[48] to test associations between perception and eating disorders risk (measured by the Eating Attitudes Test with three subscales: dieting, bulimia, and oral control, where higher scores mean higher risk).[50] Body satisfaction score (regarding one's own weight and appearance, where a higher score means higher satisfaction) was an intermediary factor (Table 1 and Appendix 2).

Among girls, increased perceived peer approval of their appearance between ages 13 and 17 was associated with increased body satisfaction, controlling for change in body mass index (BMI) over the same period (β=2·567, 95% CI 1·405, 3·729; p-value<0·0001). An increase in body satisfaction, in turn, was associated with decreased dieting risk score by age 22 (β=-0·048, 95% CI -0·088, -0·008; p-value=0·019) (Appendix Table A2.3). This translated into a statistically significant indirect association between perceived peer approval and dieting (β=-0·124, 95% CI -0·008, -0·240, p-value= 0.036), with similar trends for bulimia and attempts to control eating as measured by oral control scores (Appendix Figure A2.1), and across levels of household wealth. The direct association between perceived approval and eating disorder risk was small and not statistically significant.

Boys' body satisfaction was also influenced by perceived peer opinion, but overall risk of eating disorders was not consistently influenced, with wealth having a moderating role (Appendix Figure A2.2). For boys in lower-wealth households, increased perception of peers' approval over time was associated with a reduction in dieting scores, with a marked reversal of this association in higher-wealth households.

These results demonstrate the importance of peer-mediated body dissatisfaction in dieting behaviours in girls, and intersectionality of normative expectations with wealth in boys, perhaps reflecting broader media influences in wealthier households. Findings suggest that interventions aiming to reduce adolescents' harmful weight control behaviour should engage peer networks in challenging unhealthy norms of body appearance.

Case 3:

What children believe to be their parents' judgments of their weight, communicated through either words or actions (e.g. weight-based teasing) is associated with body dissatisfaction,[51] and has in turn been linked to adverse mental health outcomes. We examine the influence of normative pressure from parents in

Brazil, where urban culture places high value on body appearance and is accepting of weight control behaviours.[52]

The Brazil data are from a birth cohort (N=5,249) from the city of Pelotas in 1993.[53] Here, we test the role of perceived parents' opinion of adolescent boys' and girls' weight at age 11 ('thin,' 'normal,' or 'fat') as a moderator of the effect of body dissatisfaction at age 15 (feeling fatter or thinner than ideal) on mental health at age 18. Mental health was measured using the Self-Reporting Questionnaire (SRQ) screening instrument (higher score indicates worse mental health).[54] We restricted the analytic sample to girls (n=1309) and boys (n=1113) with normal BMI at age 11 so that our gender norms proxy – perceived parental opinion for boys or girls – was unlikely to reflect genuine parental health concerns about overweight or underweight status (Appendix 3).

We found that a higher percentage of normal-BMI girls than boys reported that their parents thought they were fat at age 11 (7.1% vs 5.8%), whereas more boys than girls reported that their parents thought they were thin (42.6% vs 36.9%). In sex-disaggregated regression, there was some evidence for an interaction between perceived parent's opinion about weight at age 11 and body dissatisfaction at age 15. Girls who thought they were fatter than ideal at age 15 had significantly poorer mental health at age 18 compared to those who were satisfied with their bodies, but only if, at age 11, they had reported that their parents thought they were fat (β=3·081, 95% CI 1·049, 5·114; p-value=0·003). In contrast, for girls who believed their parents thought they were normal or thin at age 11, feeling fatter than ideal at age 15 was not associated with SRQ scores (Figure A3.1). We did not observe a similar pattern among boys, suggesting that parents' opinions about body image operate differently for girls' and boys' mental health. Thus, perceived parental opinion about weight appears to be a determining factor in whether girls desiring thinness impacts their mental health.

The long-term contribution of normative parental influences to girls' later mental health in Brazil suggests a more powerful influence than previously documented. These findings further emphasise the importance

of multi-level interventions across influential groups, such as parents and teachers, to temper socially-driven health inequities.

Case study 4. School peer influences on adolescent health in the USA

Pressure to conform to restrictive gender norms can have profound effects on adolescents' mental health.[55–57] Negative social sanctions for transgressing norms are particularly salient during adolescence, when adolescents seek identity through group membership.[9,58] Sanctions can include bullying or ostracism by peers, and scolding or punishment by caretakers and/or teachers.[7] Here, we examine a pathway to risky health behaviours and poor outcomes from non-conformity with gender norms in schools.

We use data from the U.S. National Longitudinal Study of Adolescent to Adult Health (Add Health),[59] a nationally representative sample of adolescents aged 11-18 years (1994-1995) (n=20,745), randomly selected from 80 paired middle and high schools. The dataset lacks gender-specific attitude questions, but is rich in behavioural and health-related data. Following the work of Fleming et.al.,[60] we created a gender normativity measure for each student using a set of factors found to discriminate between binary sex assignment in the survey (Appendix Table A4.1). For the gender norms proxy, sex-specific individual scores were aggregated to the median of same-sex school-level peers. We tested non-conformity to dominant gender norms, expressed as the difference between an individual's estimated gender normativity and the median of their same-sex school peers, on health.

For each outcome, we conducted sex-stratified piecewise linear regressions to estimate separate effects of more typically feminine and more typically masculine behaviours compared to the median of their school, controlling for an individual's own gender normativity, birth year, race/ethnicity, and school fixed effects (Appendix Table A4.6). Standardised regression coefficients are plotted for girls (Figure 3 panel a) and boys (Figure 3 panel b) (also in Appendix Table A4.6).

Multiple health-related outcomes were associated with gender norm non-conformity. Boys and girls reporting more typically 'masculine' behaviours than their same-sex peers were significantly more likely to report risky behaviours, for example engaging in delinquent behaviour (β=0·158, 95% CI 0·015, 10·531; p-value <0·0001 for girls and β=0·399, 95% CI 0·028, 14·426; p-value <0·0001 for boys). On the other hand, boys and girls reporting more typically 'feminine' behaviours, were more likely to report weight loss behaviours (β=0·228, 95% CI 0·025, 9·265; p-value <0·0001 for girls and β=0·143, 95% CI 0·018, 7·774; p-value <0·0001 for boys). Girls were more likely to report increased depressive symptoms, and suicidal ideation and attempts with increasing difference in either direction (more typically 'masculine' or 'feminine') from peers' median gender normativity score. Results were similar controlling for household socioeconomic status (Appendix Table A4.7).

In summary, US students at the extremes of a gender-normative measure relative to other students in their school may suffer multiple health-related effects. Negative sanctions from gender-norm dominant peers may be one of the paths through which these associations operate. These results highlight the need to address stigma and negative behavioural and mental health consequences associated with gender non-conformity in schools.

Case study 5. Premarital sex and HIV status in Zambia

Sub-Saharan Africa has the highest prevalence of human immunodeficiency virus (HIV) infection globally, with new cases concentrated among adolescents[44] and disproportionately among girls.[31,61] Gender norms and power imbalances play a key role in HIV acquisition,[62–64] as they impact, for instance, condom access and use.[62,63] In the USA, embarrassment may prevent adolescents from receiving HIV information, seeking contraception, using condoms, or accessing care.[65,66]

We examine a gendered pathway to HIV infection among youth in Zambia through community expectations of appropriate sexual behaviour.[67,68] Where social norms against premarital sex exist, we

hypothesised that youth engaging in premarital sex would refrain from talking about it (with peers, parents, or health professionals), reducing their ability to learn about and access HIV protection and increasing their acquisition risk. We also hypothesised a greater impact on girls than boys, partly because of double standards[10,69] regarding appropriate sexual behaviour.

We analysed data for young women (n=1669) and men (n=1285) (ages 15-24 years) from the 2007 DHS in Zambia, one of six countries with HIV status information and balanced questions about expectations around premarital sex (Appendix 5). The gender norms proxy was adult (ages 25-49) women and men's attitudes about premarital sex, obtained by aggregating sex-specific data to 18 regional and urban-rural strata. We tested the effect of adult non-compliance with norms for premarital sex, expressed as the discordance between adult attitudes and their behaviours (believing premarital sex to be wrong, but engaging in it), on HIV acquisition risk among youth (n=2954).

Attitudes towards premarital sex did not vary substantially by sex or region in Zambia and were conservative: more than 80% of adults disapproved of premarital sex in most regions (Figure 4, panel a). In contrast, attitudes and behaviours were mostly discordant for men (most disapproved of premarital sex, but were assessed as having engaged in it, panel b), whereas women were more likely to be concordant (most disapproved of premarital sex and refrained from it). Women's perceptions of what most other women did (descriptive norms of high perceived prevalence of premarital sex) were discordant with their own behaviours (lower prevalence of premarital sex, panel c). Panel d illustrates substantial heterogeneity in HIV prevalence among youth (15-24 years) across Zambia (range 3-27%), disproportionately affecting young women in urban regions.

At the regional level, an increasing proportion of adult women (25-49 years) who refrained from engaging in premarital sex was associated with reduced HIV prevalence among adolescent women (Pearson correlation, rho=-0·43; p-value=0·077), while conservative attitudes were not. Importantly, discordance among adult women was strongly correlated with adolescent women's HIV prevalence (rho=0·63; p-

value=0·005), explaining an additional 20% of the variation in adolescent women's HIV status over behaviour alone. Furthermore, in sex-stratified Poisson regressions, we found that a 10% increase in discordance among adult women or adult men was associated with a 27% (RR=1·27, 95% CI 1·11, 1·45; p-value=0·001) or 28% (RR=1·28; 1·05, 1·56; p-value=0·015) increase, respectively, in individual-level relative risk of HIV for adolescent women, controlling for demographic and regional-level factors (Appendix 5). Risks were similar for adolescent men, but not statistically significant.

These results illustrate that gender norm non-compliance can harm health, here the risk of HIV infection, with potentially fatal consequences. Given sexual double standards,[10,69] young women may especially avoid seeking information, negotiating condom use, or seeking care to minimise risks of premarital sex, as they may anticipate heightened disapproval, relative to men. Efforts to protect women from harm associated with sexual activity should consider the normative environment in which adolescents' sexual relationships take place.

Case study 6. Women working outside the home and intimate partner violence in Nigeria

Gender norms intersect with power as adolescents move into early adulthood,[5,7,8,43,70] with unequal power relations shaping and being shaped by gender inequalities and restrictive gender norms.[10,13] Those in power benefit from, and seek to uphold, the existing social order by (consciously or unconsciously) sanctioning non-compliers.[71,72] We examine a pathway through which gendered power disparities can generate punishment (privately, at home) for women who violate the gender order by working outside the home.

Evidence is mixed on whether female labour force participation (FLFP) increases[73–76] or reduces[77,78] women's risk of intimate partner violence (IPV) in low gender-equality contexts, as IPV largely takes place in private. FLFP can be protective for working women in countries where most women work, but may be a risk factor for IPV in countries where most women do not.[78,79] We tested whether women who

work outside the home are at increased IPV risk relative to women who do not in two types of communities in Nigeria: communities where few women work outside the home and communities where FLFP is more normative.

We used data from the 2014 cluster sample design Violence against Children Survey (VACS) on experience of IPV for female youth (n=1,633, ages 13-24) (Appendix 6). FLFP was based on self-reported work outside the home in the last week. We used intraclass correlation coefficients (ICC) to detect that FLFP was clustered at the community level for girls (but not boys), with sufficient heterogeneity across communities to test our hypothesis. Assuming equal economic opportunities for work across communities, a low proportion of young women engaging in work outside the home was our gender norms proxy reflecting restrictive norms around women's mobility and opportunities to earn income. Communities were then classified as either: 1) FLFP-high (assumed absence of restrictive norms around FLFP), or 2) FLFP-low (assumed presence of norms sanctioning FLFP), based on a data-driven cut-point of 28% of female respondents engaging in outside labour. Results were robust to different cut-points (data not shown).

There were no statistically significant differences in overall past-year exposure to sexual or physical IPV for all women between the two community types (adjusted Wald tests [FLFP-high 7·3% (1·16); FLFP-low 7·9% (1·50); p-value=0·733]). Using logistic regression controlled for age, marital status, and having ever attended school, we found that women who worked in FLFP-low communities had significantly higher odds of experiencing past-year IPV compared to non-working women [OR=2·381, 95% CI 1·292, 4·389; p-value=0·006]. However, in FLFP-high communities, women's IPV risk did not differ by working status (Appendix Table A6.4).

The increased risk of IPV exposure for working women in FLFP-low communities suggests that some male partners may use IPV to punish women for transgressing gender norms around work and the perceived threat to their masculine role as breadwinner or power-holder. Although early transgressors of

restrictive norms may experience IPV as a consequence, they may also initiate long-term norm changes in ways that improve employment opportunities and health for future generations.[80] We examine elsewhere the implications of gender norms for FLFP and women's health across geo-cultural contexts[81] and time.[82]

These findings have important implications for interventions at the intersection of gender equality and global health and development—for example, efforts to empower women through employment or micro-finance of small businesses. When instituting such empowerment programmes, risks of harm to those encouraged to challenge restrictive gender norms must be anticipated, and harm prevention and mitigation strategies implemented for effective reduction in gender inequalities and health inequities.

Opportunities and challenges

Our case studies provided practical opportunities to conduct gender norm-health research using existing survey data in new ways. For example, geospatial clustering in case 1 revealed regional variation in gender norms where sex intersected with religious identity to produce large inequities in healthcare seeking – a finding that individual-level analyses might miss. Clustering communities together overcame the challenge of small numbers (i.e., precision) when estimating group-level behaviours for communities with few sick children. This innovative approach to identifying gender inequities could be extended to other health-related indicators and countries.

The inclusion of a targeted question in case studies 2 and 3 about 'what adolescents thought that others thought' was useful for estimating the normative influence of peers and parents. Similarly targeted questions could be added with limited additional expense to future surveys. In case 4, the construction of a gender normativity index enabled the use of a dataset rich in measures of gender-related behaviours to study gender non-conformity and health. This novel approach could be generalised to datasets such as the Global school-based student health survey to expand this exploration in diverse contexts.

The measure of discordance between group-level attitudes and behaviours related to premarital sex in case 5 disrupted the common practice of using only attitudes or only behaviours as gender norm proxies. Contrasting other matched attitude-behaviour pairs in this way could generate additional new insights for gendered pathways to health, as shown here for the acquisition of HIV. Finally, case 6 demonstrates how ICC, which is traditionally used to estimate effective sample size in clustered study designs, can be reinterpreted to identify sufficient clustering of behaviours to study within-country variation in gender norms.

Nevertheless, we encountered multiple data limitations, not the least of which was relying on sex-disaggregated data to study gender. In recent decades, global health leaders have increasingly recommended incorporation of gender in data systems.[12,83–89] A comprehensive United Nations report on gender statistics recommended that data should systematically be sex-stratified; measure gender facets, including norms and relations; reflect the diversity of women and men, capturing multi-dimensional aspects of their lives; and be free of gender stereotypes and biases.[88] While these guidelines provide a useful framework for collecting gender-sensitive data, none of the 17 publicly available data sources we explored (Appendix Table A8.1) were designed accordingly. The substitution of a binary sex indicator for gender in sex-disaggregated data represents a missed opportunity to study gender and health along a continuum of experiences and may have introduced important misclassification biases in our analyses.

Moreover, many datasets lacked the combination of gender-related attitudes or behaviours and health outcomes required for understanding pathways between them. Even when both were available, data were often missing for certain demographic groups or regions of the world. For example, DHS represent low- and middle-income countries and data were often missing for men (e.g., questions on child care), women (e.g., questions on some sexual practices), or certain age groups (e.g., children 6-14 years and women over 49), which can bias data interpretation. In some cases, the available proxy was perhaps too distal

from the health outcome of interest, or confounded by intermediate factors, to detect an association (e.g., between attitudes around IPV and childhood malnutrition).[90]

Additional data limitations included the inability to stratify samples by subgroups, both because of lack of indicators (e.g., missing race/ethnicity information) and small samples. Attempts to disaggregate national survey data to sub-national levels or across socio-economic strata decreased statistical power, limiting our capacity to study impacts of intersecting disadvantage with precision.

Notably, we encountered survey questions that belied gender-biased assumptions in their construction. For example, we used the rich attitudinal data in the World Values Survey (WVS) to explore adult self-rated health and gender norms around employment. However, the employment status question cannot account for cross-cultural differences in the meaning of self-employment, and includes the gender-biased term "housewife" as one of its English-version response categories. Forty-three of 46 surveys back-translated to English used a housewife-like phrase or word (21 of 24 languages and 33 of 36 countries) as opposed to a gender-neutral description (Appendix Table A7.3). Such variation made the category unreliable for cross-national comparisons and likely biased. Additionally, phrasing of attitudinal questions, such as "Pre-school children suffer with a working mother," communicates the stereotype that mother's role is at home as caregiver while father's employment-related absence is inconsequential for young children. It is also unclear whether the question refers to a situation where both parents work, or only the mother versus the father works. Furthermore, questions phrased with the terms "wife" or "husband" suggest that the questions only apply to married couples in heterosexual unions.

Finally, women and men may answer survey questions based on gendered expectations of what they think they should say rather than on their lived experiences, particularly around such gender-charged topics as sexual behaviour or eating disorders. Potentially biased responses may have led us to reproduce current, potentially biased understandings of gendered behaviour and health risk, while missing important at-risk groups.

Combined, these data limitations hindered our exploration of how, and by whom, norms are enforced and the differential impacts of norm violations across the life course and world regions. Heise, Greene et al. argued that gender "biases can be manifested and reinforced by research methodologies".[10] While publicly available survey data provided many opportunities for testing hypotheses about gender norms and health, care is required to avoid introducing or perpetuating bias when constructing and using gender norms proxies from these data.

Research agenda

In future research, we join many others in advocating for collecting survey-based data on all facets of gender, including data for gender minorities.[12,83–89] We also advocate for balanced survey data in which men and women are equally represented across age groups and asked the same unbiased attitudinal and behavioural questions, enabling gender-comparative research. Given constrained resources, we recognise that choices must be made in designing surveys, but each confers trade-offs that should be analysed from an intersectional lens encompassing gender. If certain domains are assumed unimportant (e.g., childcare provided by men) and hence not measured, then we will not be able to assess or effect change.[91] Data that reflect society not only as it is, but also as we aspire for it to be, are critical for monitoring progress on SDGs. Identifying and better measuring current and evolving gender norms across cultures, life stages, and areas of society will enable more robust study of gender norms and health.

In addition to more gender-sensitive data, we require more research on gendered pathways to health, including integrating qualitative research to unpack the origins, preservation, and shifts in gender norms. The collection of harmonised and consistent data across contexts and over time (e.g., standards for measuring gender and gender norms across global surveys), combined with longitudinal methods, would allow for cross-national comparisons, assessments of cohort effects and causal impact, and monitoring of gender norm evolution. Methods that overlay different types of data, such as survey-based and geospatial data, could utilise external factors (e.g., climate change and economic shocks) to identify locations of

gender-based discrimination. Machine learning algorithms and natural language processing could offer novel approaches to eliminating gender-related biases coded in large existing datasets.

Finally, we advocate for enhanced collaborations across the humanities and social and health sciences to provide conceptual bridges for effective data use around an evidence-based research agenda. Representation from domain experts and gender scholars, survey designers and analysts, and community partners and policy makers will allow for data systems that enable studying health at the intersection of gender and other social determinants (e.g., race, religion, and social class). Identifying mechanisms for safely sharing and analysing survey datasets is critical for safeguarding privacy while enabling new opportunities to study this intersectionality in global health research.

Conclusion

A variety of analytic tools applied to existing survey-based data across six case studies examined how restrictive gender norms can harm the health of women and men, boys and girls, across diverse settings and outcomes. We demonstrated how to construct creative gender norm proxies and conduct analyses using a variety of methods to gain novel insights into links between gender norms and health using available survey data. We also presented key limitations to advancing the field.

Four key findings emerged that have important implications for programmatic practice and policy. First, as the case study on care-seeking for childhood illness in Ethiopia shows, gender norms may intersect with other social determinants to impact health, sometimes in unexpected ways, deviating from what practitioners and policy-makers might intuitively anticipate. Second, as evidence in Brazil and South Africa suggests, early gender normative influences may affect health in different ways for boys and girls, and differentially by family context. Third, as the Add Health data in the US and the VACS data in Nigeria highlight, gender non-conformity and norm transgression may be harmful to health, particularly when challenging power relations and triggering negative sanctions. Finally, as shown with proxy

measures across case studies, the impact of gender norms can be highly context-specific. Therefore, generalisations around gender norms can be counterproductive, misleading, or even harmful. Ecological studies (e.g., with national indicators of gender inequality), while informative for hypothesis generation, belie the complexity and importance of local factors that influence relationships between gender norms and health. A deep understanding of sociocultural contexts, aided by qualitative research, is required to design effective prevention and mitigation strategies for socially-driven health inequities, and ongoing monitoring must be in place to identify, support and protect those who challenge restrictive gender norms and existing gender-based power differentials. Public health programs and policies that are locally relevant while globally active are central to achieving both gender equality and health. Progress can be accelerated through improved qualitative and quantitative data collection, analysis, and interpretation that accounts for the pervasive role of gender norms in shaping human health and well-being.

553 **Author Contributions**

554 AW, BC and VM worked closely with analysts and data owners to conceive, plan, and interpret results
555 from the case studies. They framed, drafted, and revised the manuscript.

556 GLD was the Principal Investigator of *The Lancet* Series on Gender and Health project and implemented
557 the multiple data contributor/partnership for the case studies. He also worked closely with the analysts
558 and data owners to conceive, plan, and interpret results from the case studies, as well as providing critical
559 input on framing, review and edits to the manuscript.

560 The following authors worked on one or more case studies, including data analyses and writing methods
561 and results (e.g., for appendices), contributed to the interpretation of the results for case studies, as well as
562 a critical review of the manuscript: for case study 1: PL (primary) with AJDB, IMG, EH, and SA, and the
563 support of working group members in Pelotas, Brazil; case study 2: SA (primary) with LR and SAN; case
564 study 3: SA (primary) with CGV and RB; case study 4: BD (primary) with JN, HB, and SA; case study 5:
565 IMG (primary) with EH and PL; case study 6: IS and LS (primary) with EH and IMG.

566 Additional input for case studies received from TN, NH, SC, and KM, as well as review and edits to the
567 manuscript.

568 DB performed literature searches for individual case studies and the overall paper and contributed to the
569 writing and review of the manuscript.

570 RG performed the analyses for the survey question translation example, prepared the data and created the
571 word clouds for the DHS modules, and contributed to the methods documentation.

572 MC provided critical input on framing, review and edits to the manuscript.

573

574 **Declaration of Interests**

575 The authors declare they have no conflicts of interest. The views expressed are those of the authors and
576 are not necessarily those of the Bill & Melinda Gates Foundation and the United Arab Emirates. As
577 corresponding author, AMW states that she had full access to all data and final responsibility to submit
578 for publication.

579

Acknowledgments

We thank our working group members[a] for their contributions to the six case studies, as well as our collaborators who worked in parallel on case studies not included in this paper. We also thank the members of the Gender Equality, Norms and Health Steering Committee[b] for their insights and critical review throughout the development of the paper.

[a]Working Group

Kathleen M. Harris, PhD, University of North Carolina at Chapel Hill

Carolina Coll, PhD, Federal University of Pelotas, Brazil

Janaina Calu Costa, MS, Federal University of Pelotas, Brazil

Leonardo Ferreira, MS, Federal University of Pelotas, Brazil

Franciele Hellwig, Federal University of Pelotas, Brazil

Eran Bendavid, MD, Stanford University

Marcella Alsan, MD, PhD, Stanford University

Sarah Hawkes, MD, PhD, University College London

James Zou, PhD, Stanford University

Londa Schiebinger,PhD, Stanford University

Lori Heise, PhD, Johns Hopkins Bloomberg School of Public Health

[b]Gender Equality, Norms and Health Steering Committee:

Gary Darmstadt, MD (Chair); Margaret Greene, PhD; Sarah Hawkes, PhD; Lori Heise, PhD; Sarah Henry, MPH; Jody Heymann, PhD; Jeni Klugman, PhD; Ruth Levine, PhD; Anita Raj, PhD; Geeta Rao Gupta, PhD

References

1 Magar V. Gender, health and the Sustainable Development Goals. *Bulletin of the World Health Organization* 2015; **93**: 743–743.

2 UN Women. Turning promises into action: Gender equality in the 2030 Agenda for Sustainable Development. 2018 http://www.unwomen.org/en/digital-library/sdg-report (accessed Oct 1, 2018).

3 Connell R. Gender, health and theory: Conceptualizing the issue, in local and world perspective. *Social Science & Medicine* 2012; **74**: 1675–83.

4 Langer A, Meleis A, Knaul FM, *et al.* Women and health: The key for sustainable development. *The Lancet* 2015; **386**: 1165–210.

5 Munoz Boudet, Ana Maria, Patetsch, Patti, Turk, Carolyn, Thumala, Angelica. On norms and agency: Conversations about gender equality with women and men in 20 countries. 2012.

6 Sen G, Östlin P. Gender inequity in health: why it exists and how we can change it. 2008.

7 Basu S, Zuo X, Lou C, Acharya R, Lundgren R. Learning to be gendered: Gender socialization in early adolescence among urban poor in delhi, india, and shanghai, china. *Journal of Adolescent Health* 2017; **61**: S24–9.

8 Lundgren R, Burgess S, Chantelois H, Oregede S, Kerner B, Kågesten AE. Processing gender: Lived experiences of reproducing and transforming gender norms over the life course of young people in Northern Uganda. *Culture, Health & Sexuality* 2018; : 1–17.

9 Kågesten A, Gibbs S, Blum RW, *et al.* Understanding factors that shape gender attitudes in early adolescence globally: a mixed-methods systematic review. *PLOS ONE* 2016; **11**: e0157805.

10 Heise L, Greene, M, Opper N, et al. Gender inequality and restrictive gender norms: Framing the challenges to health. *Lancet* 2019.

11 Krieger N. Genders, sexes, and health: What are the connections—and why does it matter? *International Journal of Epidemiology* 2003; **32**: 652–7.

12 Schiebinger, L., Klinge, I., Sánchez de Madariaga, I., Paik, H. Y., Schraudner, M., and Stefanick, M. (Eds.) (2011-2015). Analyzing how sex and gender interact | Gendered Innovations. http://genderedinnovations.stanford.edu/methods/how.html (accessed Oct 1, 2018).

13 Sen G, Östlin P, George, A. Unequal, unfair, ineffective and inefficient: Gender inequity in health—why it exists and how we can change it. *Global Public Health* 2008; **3**: 1–12.

14 Springer KW, Mager Stellman J, Jordan-Young RM. Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health. *Social Science & Medicine* 2012; **74**: 1817–24.

15 Ahmed T, Vafaei A, Auais M, Guralnik J, Zunzunegui MV. Gender roles and physical function in older adults: Cross-sectional analysis of the International Mobility in Aging Study (IMIAS). *PLoS ONE* 2016; **11**: e0156828.

16 Cullen MR, Baiocchi M, Eggleston K, Loftus P, Fuchs V. The weaker sex? Vulnerable men and women's resilience to socio-economic disadvantage. *SSM - Population Health* 2016; **2**: 512–24.

17 Hosseinpoor AR, Stewart Williams J, Amin A, *et al.* Social determinants of self-reported health in women and men: Understanding the role of gender in population health. *PLoS ONE* 2012; **7**: e34799.

18 Mokdad AH, Forouzanfar MH, Daoud F, *et al.* Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2016; **387**: 2383–401.

19 Institute for Health Metrics and Evaluation. Global health data exchange: Global Burden of Disease (GBD) Results Tool. 2018. http://ghdx.healthdata.org/gbd-results-tool (accessed Oct 1, 2018).

20 Snow RC. Sex, gender, and vulnerability. *Glob Public Health* 2008; **3 Suppl 1**: 58–74.

21 World Health Organization. Gender and road traffic injuries. 2002.

22 Messing K, Östlin P, World Health Organization. Gender equality, work and health: A review of the evidence. Geneva: World Health Organization, 2006.

23 Courtenay WH. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. *Social science & medicine* 2000; **50**: 1385–401.

24 Fleming PJ, Agnew-Brune C. Current trends in the study of gender norms and health behaviors. *Current opinion in psychology* 2015; **5**: 72–7.

25 Lazar M. Feminist critical discourse analysis: Gender, power and ideology in discourse. Springer, 2005.

26 Phillips SP. Defining and measuring gender: A social determinant of health whose time has come. *International Journal for Equity in Health* 2005; **4**. DOI:10.1186/1475-9276-4-11.

27 Morgan R, George A, Ssali S, Hawkins K, Molyneux S, Theobald S. How to do (or not to do)… gender analysis in health systems research. *Health Policy and Planning* 2016; **31**: 1069–78.

28 Das Gupta M, Zhenghua J, Bohua L, Zhenming X, Chung W, Hwa-Ok B. Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *Journal of Development Studies* 2003; **40**: 153–87.

29 Diamond-Smith N, Luke N, McGarvey S. 'Too many girls, too much dowry': Son preference and daughter aversion in rural Tamil Nadu, India. *Cult Health Sex* 2008; **10**: 697–708.

30 Hesketh T. Selecting sex: the effect of preferring sons. *Early Hum Dev* 2011; **87**: 759–61.

31 Jayachandran S. The roots of gender inequality in developing countries. *Annual Review of Economics* 2015; **7**: 63–88.

32 Calu Costa J, Wehrmeister FC, Barros AJ, Victora CG. Gender bias in careseeking practices in 57 low- and middle-income countries. *J Glob Health* 2017; **7**: 010418.

669 33 Bohn JA, Kassaye BM, Record D, *et al.* Demographic and mortality analysis of hospitalized children at
670 a referral hospital in Addis Ababa, Ethiopia. *BMC Pediatr* 2016; **16**: 168.

671 34 Gelaw YA, Biks GA, Alene KA. Effect of residence on mothers' health care seeking behavior for
672 common childhood illness in Northwest Ethiopia: A community based comparative cross-sectional
673 study. *BMC Res Notes* 2014; **7**: 705.

674 35 Getis A, Ord JK. The analysis of spatial association by use of distance statistics. In: Perspectives on
675 Spatial Data Analysis. Springer, 2010: 127–45.

676 36 Sen G, Iyer A. Who gains, who loses and how: Leveraging gender and class intersections to secure
677 health entitlements. *Soc Sci Med* 2012; **74**: 1802–11.

678 37 Hopcroft RL. Parental status and differential investment in sons and daughters: Trivers-Willard
679 revisited. *Social Forces* 2005; **83**: 1111–36.

680 38 Veller C, Haig D, Nowak MA. The Trivers-Willard hypothesis: Sex ratio or investment? *Proc Biol Sci*
681 2016; **283**. DOI:10.1098/rspb.2016.0126.

682 39 Tenenbaum HR, Leaper C. Are parents' gender schemas related to their children's gender-related
683 cognitions? A meta-analysis. *Dev Psychol* 2002; **38**: 615–30.

684 40 van der Pol LD, Groeneveld MG, van Berkel SR, *et al.* Fathers' and mothers' emotion talk with their
685 girls and boys from toddlerhood to preschool age. *Emotion* 2015; **15**: 854–64.

686 41 Yu C, Zuo X, Blum RW, *et al.* Marching to a different drummer: A cross-cultural comparison of young
687 adolescents who challenge gender norms. *J Adolesc Health* 2017; **61**: S48–54.

688 42 De Meyer S, Kågesten A, Mmari K, *et al.* 'Boys should have the courage to ask a girl out': Gender
689 norms in early adolescent romantic relationships. *J Adolesc Health* 2017; **61**: S42–7.

690 43 Vu L, Pulerwitz J, Burnett-Zieman B, Banura C, Okal J, Yam E. Inequitable gender norms from early
691 adolescence to young adulthood in Uganda: Tool validation and differences across age groups. *J*
692 *Adolesc Health* 2017; **60**: S15–21.

693 44 Patton GC, Sawyer SM, Santelli JS, *et al.* Our future: A Lancet commission on adolescent health and
694 wellbeing. *Lancet* 2016; **387**: 2423–78.

695 45 Ricciardelli LA, McCabe MP, Holt KE, Finemore J. A biopsychosocial model for understanding body
696 image and body change strategies among children. *Journal of Applied Developmental Psychology*
697 2003; **24**: 475–95.

698 46 Bair A, Steele JR, Mills JS. Do these norms make me look fat? The effect of exposure to others' body
699 preferences on personal body ideals. *Body Image* 2014; **11**: 275–81.

700 47 Brady J, Iwamoto DK, Grivel M, Kaya A, Clinton L. A systematic review of the salient role of feminine
701 norms on substance use among women. *Addict Behav* 2016; **62**: 83–90.

48 Richter L, Norris S, Pettifor J, Yach D, Cameron N. Cohort profile: Mandela's children: The 1990 birth to twenty study in South Africa. *International Journal of Epidemiology* 2007; **36**: 504–11.

49 Szabo CP, Berk M, Tlou E, Allwood CW. Eating disorders in black South African females. *South African Medical Journal* 1995; **85**. https://www.ajol.info/index.php/samj/article/view/156125 (accessed Oct 1, 2018).

50 Garner DM, Olmsted MP, Bohr Y, Garfinkel PE. The eating attitudes test: Psychometric features and clinical correlates. *Psychol Med* 1982; **12**: 871–8.

51 Reel J, Voelker D, Greenleaf C. Weight status and body image perceptions in adolescents: current perspectives. *Adolescent Health, Medicine and Therapeutics* 2015; : 149.

52 Goldenberg M. The body as capital: Understanding Brazilian culture. *Vibrant - Vibrant Virtual Brazilian Anthropology* 2010; **7**: 220–38.

53 Victora CG, Hallal PC, Araújo CL, Menezes AM, Wells JC, Barros FC. Cohort profile: The 1993 Pelotas (Brazil) birth cohort study. *Int J Epidemiol* 2008; **37**: 704–9.

54 Beusenberg M, Orley JH, Health WHOD of M. A user's guide to the Self-Reporting Questionnaire (SRQ). Geneva: Geneva : World Health Organization, 1994 http://apps.who.int/iris/handle/10665/61113 (accessed Oct 1, 2018).

55 Gordon AR, Krieger N, Okechukwu CA, *et al.* Decrements in health-related quality of life associated with gender nonconformity among U.S. adolescents and young adults. *Qual Life Res* 2017; **26**: 2129–38.

56 Roberts AL, Rosario M, Slopen N, Calzo JP, Austin SB. Childhood gender nonconformity, bullying victimization, and depressive symptoms across adolescence and early adulthood: An 11-year longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry* 2013; **52**: 143–52.

57 Valentine SE, Shipherd JC. A systematic review of social stress and mental health among transgender and gender non-conforming people in the United States. *Clin Psychol Rev* 2018; published online March 28. DOI:10.1016/j.cpr.2018.03.003.

58 van de Bongardt D, Reitz E, Sandfort T, Deković M. A meta-analysis of the relations between three types of peer norms and adolescent sexual behavior. *Personality and Social Psychology Review* 2015; **19**: 203–34.

59 Harris KM. The Add Health study: Design and accomplishments. 2013. DOI:10.17615/c6tw87.

60 Fleming PJ, Harris KM, Halpern CT. Description and Evaluation of a Measurement Technique for Assessment of Performing Gender. *Sex Roles* 2017; **76**: 731–46.

61 Sia D, Onadja Y, Nandi A, Foro A, Brewer T. What lies behind gender inequalities in HIV/AIDS in sub-Saharan African countries: Evidence from Kenya, Lesotho and Tanzania. *Health Policy Plan* 2014; **29**: 938–49.

737 62 Gottert A, Barrington C, Pettifor A, *et al.* Measuring men's gender norms and gender role
738 conflict/stress in a high HIV-prevalence South African setting. *AIDS Behav* 2016; **20**: 1785–95.

739 63 Haberland. The case for addressing gender and power in sexuality and HIV education: A
740 comprehensive review of evaluation studies. *International Perspectives on Sexual and Reproductive*
741 *Health* 2015; **41**: 31.

742 64 Wingood GM, DiClemente RJ. Application of the theory of gender and power to examine HIV-related
743 exposures, risk factors, and effective interventions for women. *Health Education & Behavior* 2000;
744 **27**: 539–65.

745 65 Commendador KA. Parental influences on adolescent decision making and contraceptive use.
746 *Pediatric Nursing* 2010; **36**: 147–56.

747 66 Daley AM. What influences adolescents' contraceptive decision-making? A meta-ethnography. *J*
748 *Pediatr Nurs* 2014; **29**: 614–32.

749 67 Exavery A, Mubyazi GM, Rugemalila J, *et al.* Acceptability of condom promotion and distribution
750 among 10-19 year-old adolescents in Mpwapwa and Mbeya rural districts, Tanzania. *BMC Public*
751 *Health* 2012; **12**: 569.

752 68 Fearon E, Wiggins RD, Pettifor AE, Hargreaves JR. Is the sexual behaviour of young people in sub-
753 Saharan Africa influenced by their peers? A systematic review. *Social Science & Medicine* 2015; **146**:
754 62–74.

755 69 Emmerink PMJ, Vanwesenbeeck I, van den Eijnden RJJM, Ter Bogt TFM. Psychosexual correlates of
756 sexual double standard endorsement in adolescent sexuality. *J Sex Res* 2016; **53**: 286–97.

757 70 Madiba S, Ngwenya N. Cultural practices, gender inequality and inconsistent condom use increase
758 vulnerability to HIV infection: narratives from married and cohabiting women in rural communities in
759 Mpumalanga province, South Africa. *Glob Health Action* 2017; **10**: 1341597.

760 71 Cislaghi B, Heise L. Four avenues of normative influence: A research agenda for health promotion in
761 low and mid-income countries. *Health Psychology* 2018; **37**: 562–73.

762 72 Cislaghi B, Heise LL. Measuring gender-related social norms: Report of a meeting. Baltimore, MD:
763 Learning Group on Social Norms and Gender-based Violence of the London School of Hygiene and
764 Tropical Medicine, 2016.

765 73 Chin Y-M. Male backlash, bargaining, or exposure reduction? Women's working status and physical
766 spousal violence in India. *Journal of Population Economics* 2012; **25**: 175–200.

767 74 Heath R. Women's access to labor market opportunities, control of household resources, and
768 domestic violence. ; : 27.

769 75 Hynes ME, Sterk CE, Hennink M, Patel S, DePadilla L, Yount KM. Exploring gender norms, agency and
770 intimate partner violence among displaced Colombian women: A qualitative assessment. *Global*
771 *Public Health* 2016; **11**: 17–33.

76 Koenig MA, Ahmed S, Hossain MB. Women's status and domestic violence in rural Bangladesh: Individual- and community-level effects. 2003; **40**: 20.

77 Kabeer N. Conflicts over credit: Re-evaluating the empowerment potential of loans to women in rural Bangladesh. *WORLD DEVELOPMENT*; : 22.

78 Vyas S, Watts C. How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published evidence. *Journal of International Development* 2009; **21**: 577–602.

79 Heise LL, Kotsadam A. Cross-national and multilevel correlates of partner violence: An analysis of data from population-based surveys. *The Lancet Global Health* 2015; **3**: e332–40.

80 Schuler SR, Nazneen S. Does intimate partner violence decline as women's empowerment becomes normative? Perspectives of Bangladeshi women. *World Development* 2018; **101**: 284–92.

81 Hallgren E, et al. Global assessment of the relationship between gender differences in self-reported health and regional differences in female employment rate and attitudes about female labour force participation. *(in preparation)*.

82 Falconi A, et al. Shifts in gender norms in women's paid employment during the world war ii era and later life health: An application of latent transition analysis. *(in preparation)*.

83 Buvinic M, Furst-Nichols R, Koolwal G. Mapping gender data gaps. 2014.

84 Government of Canada CI of HR. How to integrate sex and gender into research. 2018; published online Feb 12. http://www.cihr-irsc.gc.ca/e/50836.html (accessed Oct 1, 2018).

85 Lawrence K, Rieder A. Methodologic and ethical ramifications of sex and gender differences in public health research. *Gend Med* 2007; **4 Suppl B**: S96-105.

86 Nieuwenhoven L, Klinge I. Scientific excellence in applying sex- and gender-sensitive methods in biomedical and health research. *J Womens Health (Larchmt)* 2010; **19**: 313–21.

87 Nowatzki N, Grant KR. Sex is not enough: The need for gender-based analysis in health research. *Health Care Women Int* 2011; **32**: 263–77.

88 United Nations Children's Fund (UNICEF). Harnessing the power of data for girls: Taking stock and looking ahead to 2030. 2016.

89 Wizemann, Theresa M; Institute of Medicine. Sex-specific reporting of scientific research: A workshop summary. Washington, D.C.: National Academies Press, 2012 DOI:10.17226/13307.

90 Chai J, Fink G, Kaaya S, *et al.* Association between intimate partner violence and poor child growth: results from 42 demographic and health surveys. *Bulletin of the World Health Organization* 2016; **94**: 331–9.

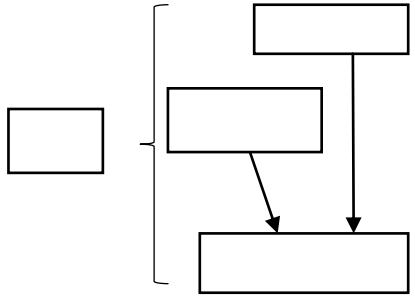
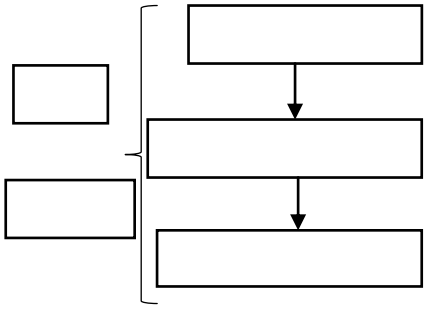
91 Hunt K, Annandale E. Just the job? Is the relationship between health and domestic and paid work gender-specific? *Sociology of Health and Illness* 1993; **15**: 632–64.

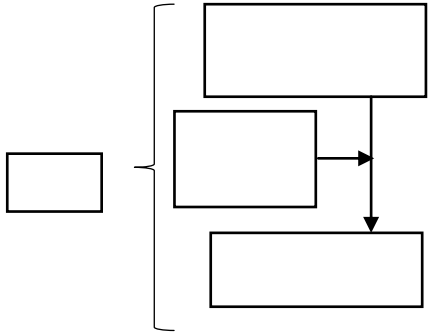
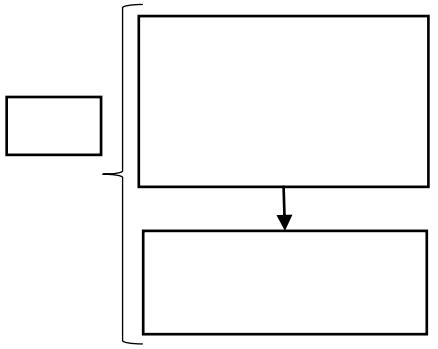
806 92 Graham AJ, Atkinson PM, Danson FM. Spatial analysis for epidemiology. *Acta Trop* 2004; **91**: 219–25.

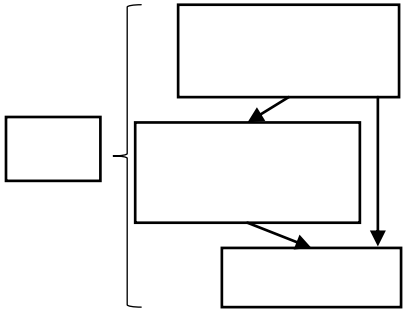
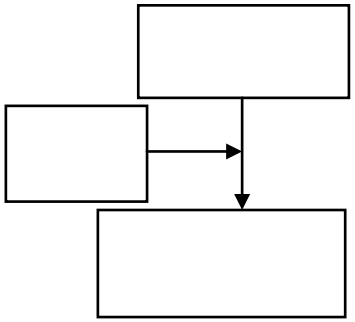
807

Tables, Figures & Panels

Table 1: Overview of case study analyses

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 1 Differential care-seeking of ill children	Gender differences in access to care	DHS, Ethiopia, 2011	Children, 0-5 y, who were ill in prior 2 weeks (n=3,161 children in 544 villages)	Indicator of communities being in a hot spot (compared to national average) for differential care-seeking for boys minus girls (proxy for gender preference)	What community factors best predict hot spots for differential care-seeking for boys vs. girls?		Differential care-seeking increased with increasing percentages of wealthy and Muslim households in communities. Differential care-seeking was greatest in communities that were both wealthy and Muslim-majority.
# 2 Community peer influence and eating disorders	Gendered health behaviours	Birth-to-20 Cohort, Soweto-Johannesburg, 1994	Male and female youth, 13-22 y (n=3273)	Individual-level perception of peers' approval of their appearance	Do adolescent perceptions of peers' opinion impact eating disorders in early adulthood? Does this vary by sex and family wealth?		As perceived peer approval increased, girls' and boys' body satisfaction increased. For girls, increasing body satisfaction was associated with a decrease in eating disorders. Boys risk of dieting varied by household wealth.

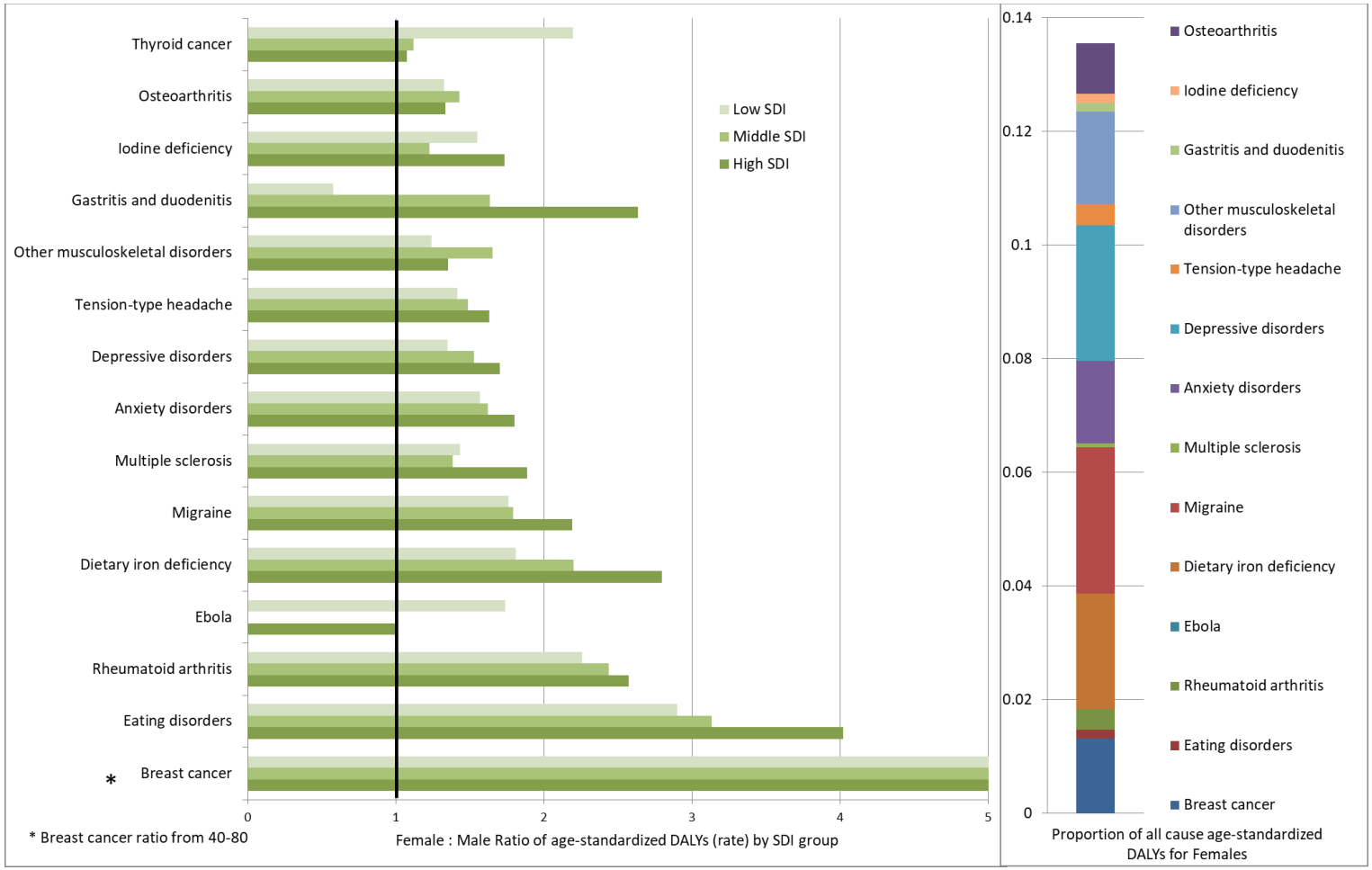
Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 3 Parental influence and mental health	Gendered health behaviours	Pelotas Birth Cohort, Brazil, 1993	Male (n=1113) and female (n=1309) youth, 11-18 y, with normal BMI at 11 y	Individual-level perception of parent's opinion of their weight	Do early adolescent perceptions of parents' opinion impact mental health in later adolescence? Does this vary by sex?		Among girls, but not boys, body dissatisfaction (feeling fatter than ideal) was associated with worse mental health outcomes when they thought their parents' opinion was also that they were fatter than ideal.
# 4 School grade peer influence and health	Gendered health behaviours	Add Health, USA, 1994-95	Male and female youth, 11-18 y (n=20,745)	Median gender normativity score of same-sex school peers (see Appendix 4 for details)	Does individual non-conformity with school peers' gender normativity impact health? Does this vary by sex and direction of non-conformity (more 'masculine' or 'feminine' than same-sex peers)?		For both girls and boys, increasing gender non-conformity with same-sex peers in either direction (i.e.: more 'masculine' or more 'feminine') was associated with increased risk for multiple health and behaviour outcomes.

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 5 Premarital sex and HIV status in Zambia	Gendered power disparities	DHS, Zambia, 2007	Female (n=1669) and male (n=1285) youth, 15-24 y, who ever had sex	Cluster-level (urban/rural region) average of male and female adult (25-49 y) attitudes about young people engaging in premarital sex.	Does community-level non-conformity with norms for premarital sex impact adolescent risk for HIV acquisition? Does this vary by sex?		In regions were most adults disapprove of premarital sex (and yet have premarital sex), sexually-active girls, but not boys, are at higher risk of positive HIV status
# 6 Female labour force participation (FLFP) and IPV in Nigeria	Gendered power disparities	VACS, Nigeria, 2014	13-24 y, females, (n=1633)	Indicator of community with a low % of women working outside the home (FLFP-low)	Does individual transgression of gender norms related to FLFP in low-FLFP communities impact a young women's risk of experiencing IPV?		Women who work outside the home experience higher rates of IPV than women who don't, but only in communities where working outside the home is not the norm.

[a] The gendered pathways provide a conceptual link to the gender system and health framework presented in Heise and Greene.[10]

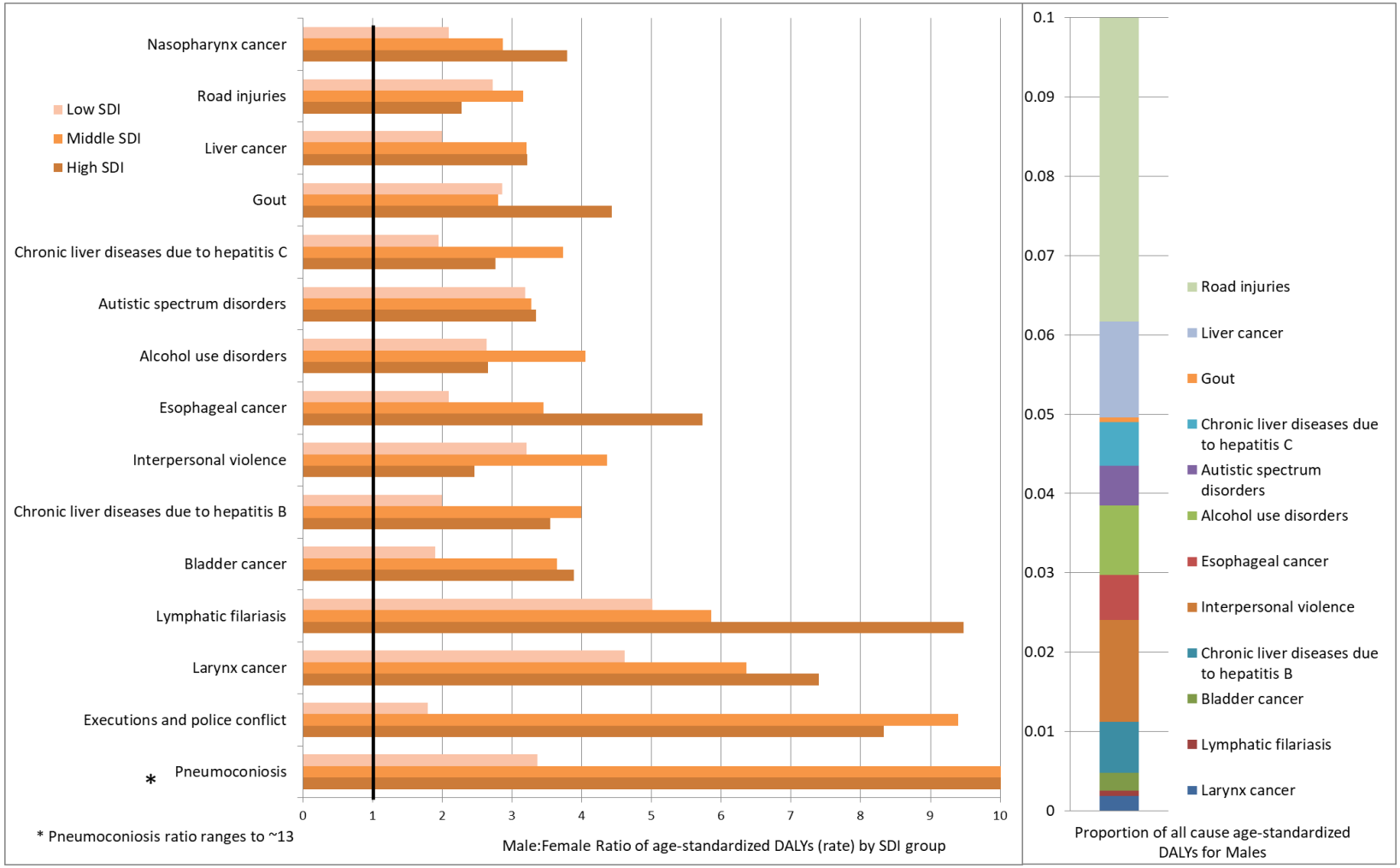
[b] The diagrams reflect the hypotheses we aimed to test and indicate a temporal causal direction. However, most of the data are cross-sectional and insufficient to determine causality.

Figure 1a: Female: Male ratio of age-standardized DALYs for low, middle, and high Sociodemographic Index (SDI)[a] groups (excluding low-middle and middle-high SDI countries for ease of data visualization)



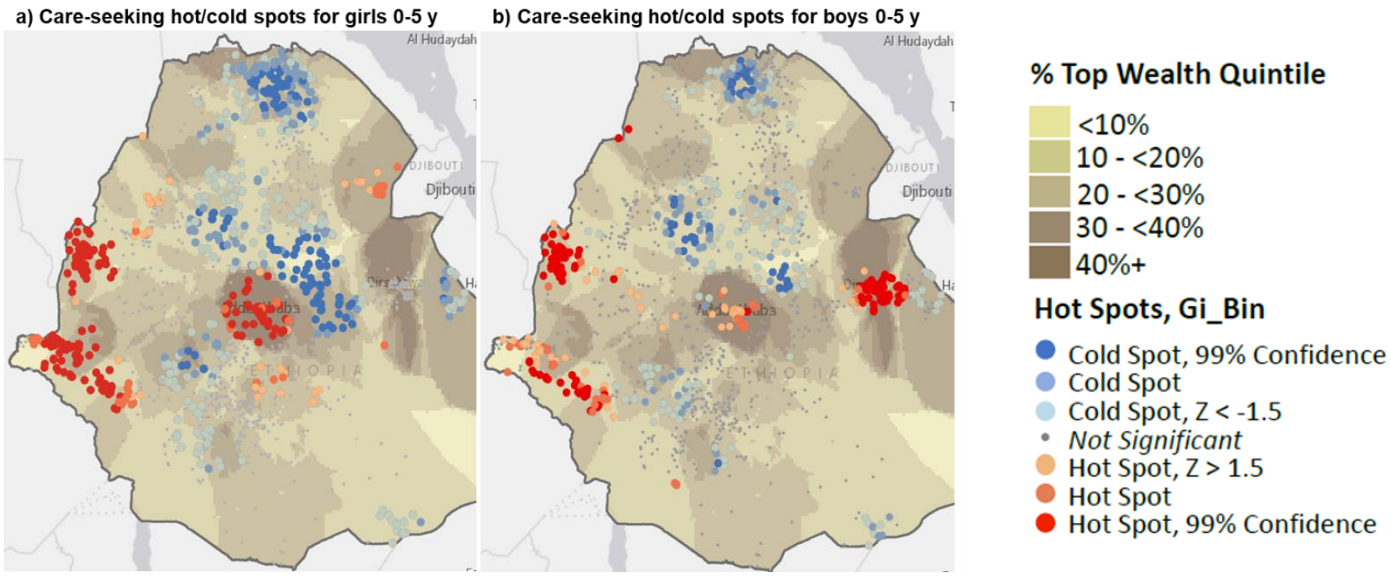
[a] SDI is comprised of: average income per person, educational attainment, and total fertility rate.

Figure 1b: Male: Female ratio of age-standardized DALYs for low, middle, and high Sociodemographic Index (SDI)[a] groups (excluding low-middle and middle-high SDI countries for ease of data visualization)



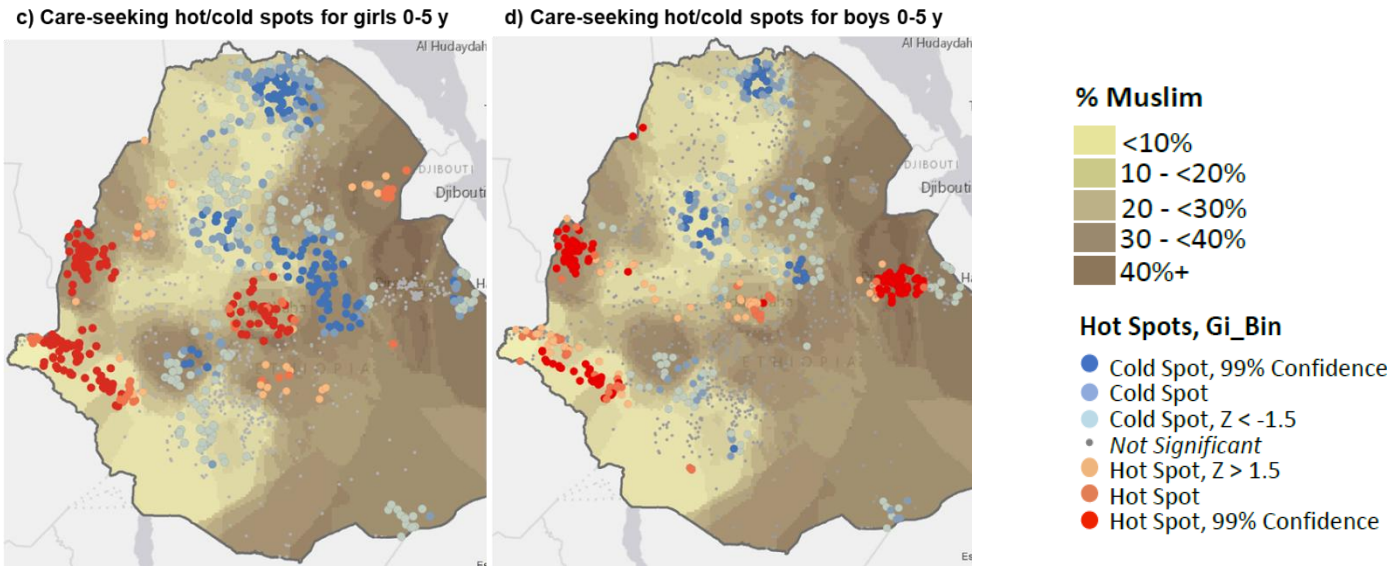
[a] SDI is comprised of: average income per person, educational attainment, and total fertility rate.

Figure 2a and 2b: Care-seeking hot/cold spots for girls (a) and boys (b) in Ethiopia by %wealthy households



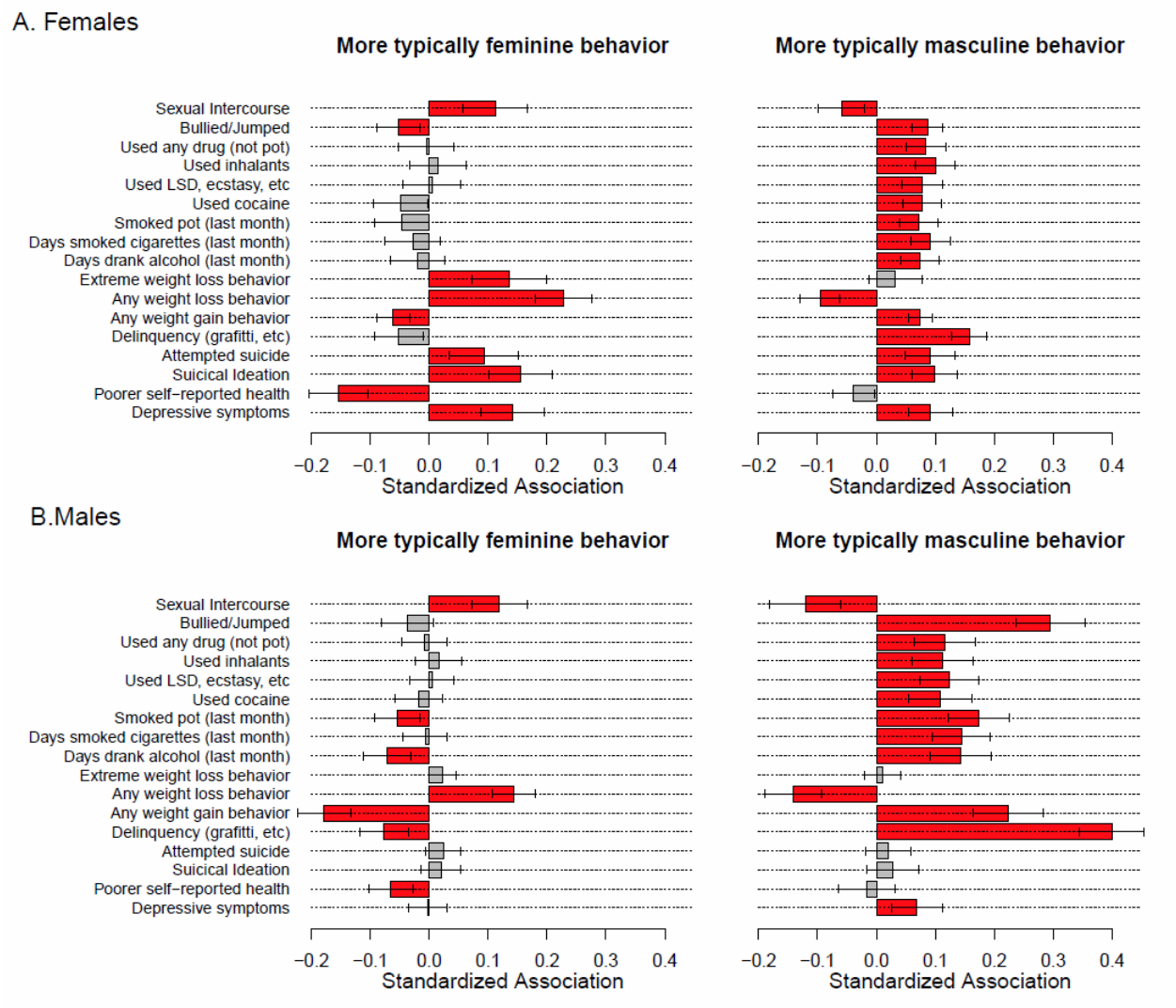
Hot spots (red) and cold spots (blue) are clusters of communities with significantly higher and lower care-seeking than the national average, respectively, for girls (a) and boys (b) separately (see Appendix 1 for details). Maps are overlaid with the spatial distribution of the percentage in the communities of top wealth quintile households (for the country). The spatial distribution is displayed using kriging, a method for interpolating spatial data.[92]

Figure 2c and 2d: Care-seeking hot/cold spots for girls (c) and boys (d) in Ethiopia by %Muslim households



Hot spots (red) and cold spots (blue) are clusters of communities with significantly higher and lower care-seeking than the national average, respectively, for girls (c) and boys (d) separately (see Appendix 1 for details). Maps are overlaid with the spatial distribution of the percentage in the communities of Muslim households. The spatial distribution is displayed using kriging, a method for interpolating spatial data.[92]

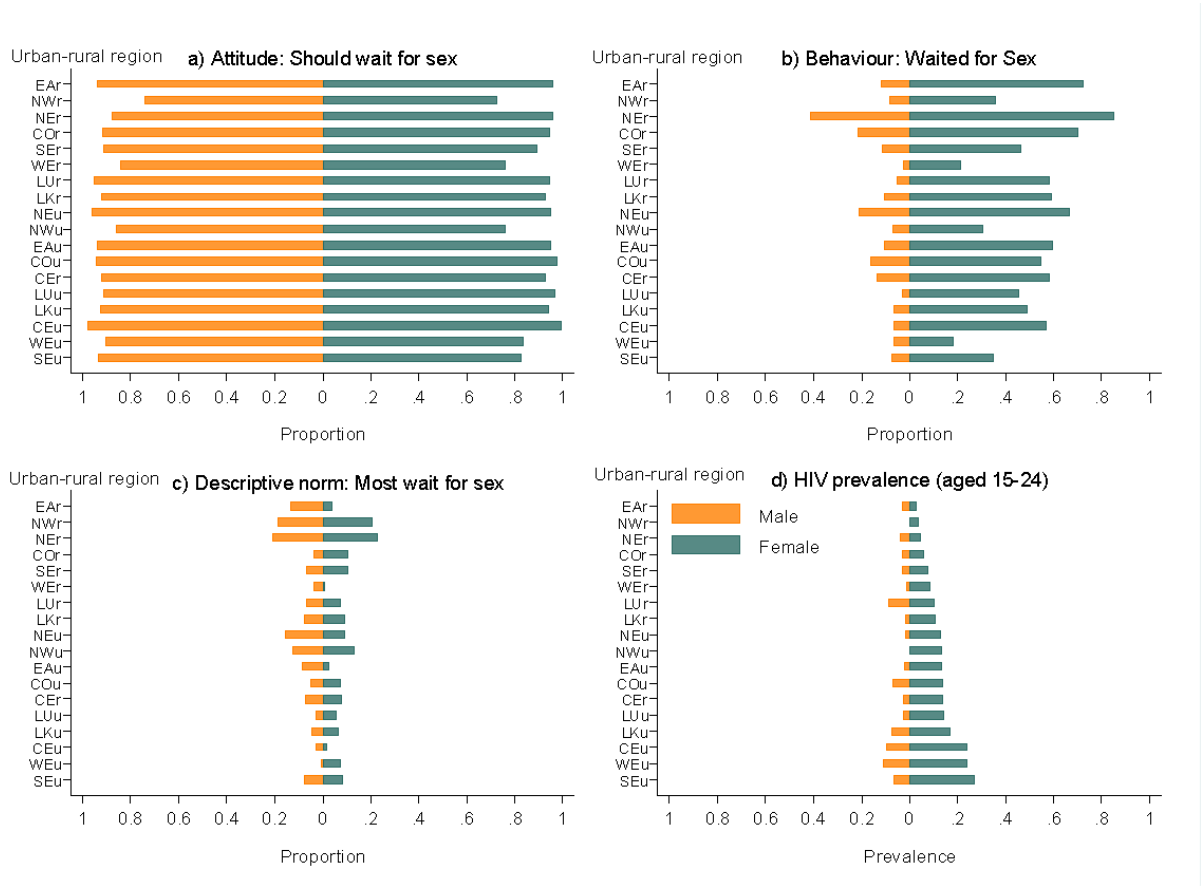
Figure 3: Estimated effects of positive and negative differences between an individual's estimated gender normativity and the median normativity of same-sex peers on health outcomes and health-related behaviours among US students, by sex.



The exposure of interest was gender norms non-conformity, or the difference between an individual's estimated gender normativity and the median of their same-sex school peers. Regressions are sex-stratified piecewise linear regressions (knot at zero) with separate effect estimates for more typically feminine and more typically masculine behaviours compared to the median of their school, controlling for an individual's own gender normativity, birth year, race, and school fixed effects. Effect estimates are

standardised so that the magnitudes can be compared across outcomes. For example, a 1 SD increase in the difference (or non-conformity) measure is associated with a 0.399 SD increase in delinquent behaviour among boys. Error bars represent 95% confidence intervals. Bars are coloured red if they are significant at the 0.01 (0.05/5) level for an appropriate Bonferroni correction based on a parallel analysis of the outcomes in the full sample, suggesting that there are 5 components.

Figure 4: Sex differentials in the proportion of adult (men and women, aged 25-49 years) for a) attitudes, b) behaviours, c) descriptive norms towards premarital sex, and d) HIV prevalence among youth (aged 15-24 years) by urban-rural regions[a] in Zambia in 2007[b]

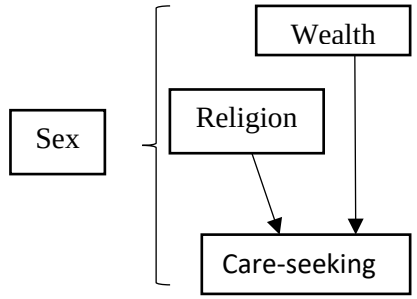
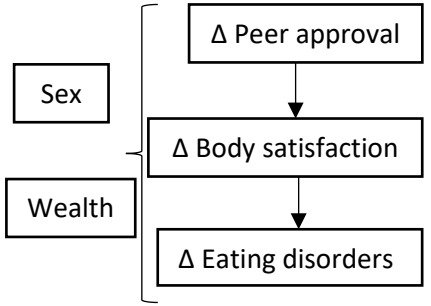


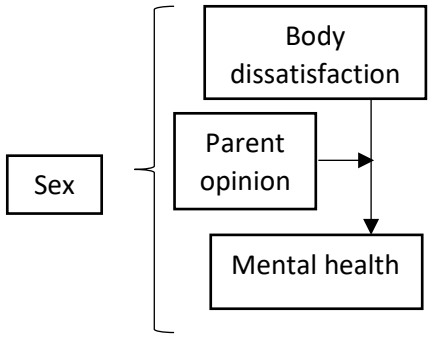
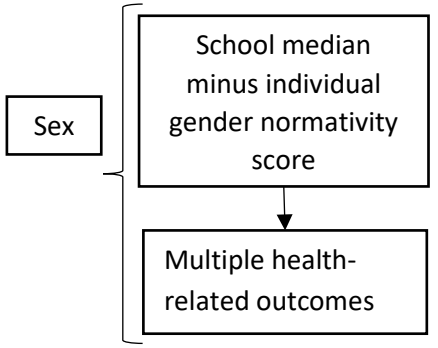
[a]Regional codes: Central "CE", Copperbelt "CO", Eastern "EA", Luapula "LU", Lusaka "LK", Northern "NE", Northwestern "NW", Southern "SE", Western "WE". The subscripts "u" and "r" stand for urban or rural region, respectively.

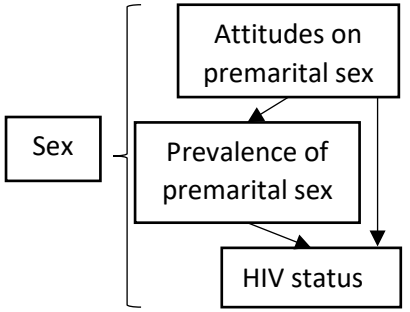
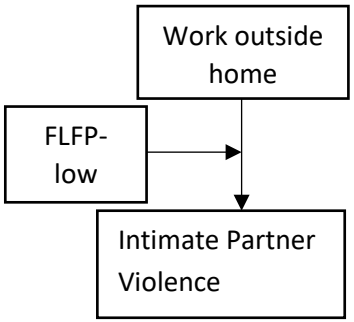
[b]Authors' estimates with information from 2007 ZDHS.

Aggregated responses were sex-stratified: men's responses about men's attitudes/behaviours and women's responses about women's attitudes/behaviours.

Table 1: Overview of case study analyses

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 1 Differential care-seeking of ill children	Gender differences in access to care	DHS, Ethiopia, 2011	Children, 0-5 y, who were ill in prior 2 weeks (n=3,161 children in 544 villages)	Indicator of communities being in a hot spot (compared to national average) for differential care-seeking for boys minus girls (proxy for gender preference)	What community factors best predict hot spots for differential care-seeking for boys vs. girls?		Differential care-seeking increased with increasing percentages of wealthy and Muslim households in communities. Differential care-seeking was greatest in communities that were both wealthy and Muslim-majority.
# 2 Community peer influence and eating disorders	Gendered health behaviours	Birth-to-20 Cohort, Soweto-Johannesburg, 1994	Male and female youth, 13-22 y (n=3273)	Individual-level perception of peers' approval of their appearance	Do adolescent perceptions of peers' opinion impact eating disorders in early adulthood? Does this vary by sex and family wealth?		As perceived peer approval increased, girls' and boys' body satisfaction increased. For girls, increasing body satisfaction was associated with a decrease in eating disorders. Boys risk of dieting varied by household wealth.

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 3 Parental influence and mental health	Gendered health behaviours	Pelotas Birth Cohort, Brazil, 1993	Male (n=1113) and female (n=1309) youth, 11-18 y, with normal BMI at 11 y	Individual-level perception of parent's opinion of their weight	Do early adolescent perceptions of parents' opinion impact mental health in later adolescence? Does this vary by sex?		Among girls, but not boys, body dissatisfaction (feeling fatter than ideal) was associated with worse mental health outcomes when they thought their parents' opinion was also that they were fatter than ideal.
# 4 School grade peer influence and health	Gendered health behaviours	Add Health, USA, 1994-95	Male and female youth, 11-18 y (n=20,745)	Median gender normativity score of same-sex school peers (see Appendix 4 for details)	Does individual non-conformity with school peers' gender normativity impact health? Does this vary by sex and direction of non-conformity (more 'masculine' or 'feminine' than same-sex peers)?		For both girls and boys, increasing gender non-conformity with same-sex peers in either direction (i.e.: more 'masculine' or more 'feminine') was associated with increased risk for multiple health and behaviour outcomes.

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 5 Premarital sex and HIV status in Zambia	Gendered power disparities	DHS, Zambia, 2007	Female (n=1669) and male (n=1285) youth, 15-24 y, who ever had sex	Cluster-level (urban/rural region) average of male and female adult (25-49 y) attitudes about young people engaging in premarital sex.	Does community-level non-conformity with norms for premarital sex impact adolescent risk for HIV acquisition? Does this vary by sex?		In regions were most adults disapprove of premarital sex (and yet have premarital sex), sexually-active girls, but not boys, are at higher risk of positive HIV status
# 6 Female labour force participation (FLFP) and IPV in Nigeria	Gendered power disparities	VACS, Nigeria, 2014	13-24 y, females, (n=1633)	Indicator of community with a low % of women working outside the home (FLFP-low)	Does individual transgression of gender norms related to FLFP in low-FLFP communities impact a young women's risk of experiencing IPV?		Women who work outside the home experience higher rates of IPV than women who don't, but only in communities where working outside the home is not the norm.

[a] The gendered pathways provide a conceptual link to the gender system and health framework presented in Heise and Greene.[10]

[b] The diagrams reflect the hypotheses we aimed to test and indicate a temporal causal direction. However, most of the data are cross-sectional and insufficient to determine causality.

*Manuscript with revisions highlighted

1

2 **Paper 2 of *The Lancet* Series on Gender Equality, Norms and Health**

3 **~~How~~ gGender norms ~~shape~~ and health: insights from global survey data**

4

5 # **Words**: Abstract: 196; body: 5,5~~66~~36

6 # **of Objects**: 5

7

Authors:

Ann Weber, PhD[1]; Beniamino Cislaghi, PhD[2]; Valerie Meausoone, BS[3]; Safa Abdalla, MD[1]; Iván Mejía-Guevara, PhD[3]; Pooja Loftus, MS[3]; Emma Hallgren, PhD [3]; Ilana Seff, MPH[4]; Lindsay Stark, DrPh[5]; ~~Devika Bhushan, MD[1];~~ Cesar Gomes Victora, MD, PhD[6]; Romina Buffarini, PhD[6]; Aluísio J. D. Barros, MD, PhD[6]; Ben Domingue, PhD[7]; Devika Bhushan, MD[1]; Ribhav Gupta, PhD[1]; Jason M. Nagata, MD[8]; Holly B. Shakya, PhD[9]; Linda Richter, PhD[10]; Shane A Norris, PhD[10]; Thoai Ngo, PhD[11]; Sophia Chae, PhD[11]; Nicole Haberland, MPH[11]; Katharine McCarthy, MPH[11]; Mark Cullen, MD[3]; Gary L. Darmstadt, MD[1,3]; for the Gender Equality, Norms and Health Steering Committee*

1 Stanford University
Department of Pediatrics, Stanford University School of Medicine
1265 Welch Road, Stanford, CA 94305 USA

2 London School of Hygiene & Tropical Medicine
15-17 Tavistock Place
London WC1H 9SH
United Kingdom

3 Center for Population Health Sciences
Stanford University School of Medicine
1070 Arastradero Rd, Palo Alto, CA 94304 USA

4 Columbia University
Department of Population and Family Health
Mailman School of Public Health
722 West 168th St. NY, NY 10032 USA

5 Brown School at Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130

6 Federal University of Pelotas, Brazil
International Center for Equity in Health
Rua Deodoro, 1120
96020-220 Pelotas, Rs, Brazil

7 Stanford University
Graduate School of Education
485 Lasuen Mall, Stanford, CA 94305 USA

8 University of California San Francisco

Department of Pediatrics
3333 California Street
San Francisco CA 94143

9 University of California San Diego
Department of Medicine, Center on Gender Equity and Health (GEH)
9500 Gilman Drive
La Jolla, CA 92093 USA

10 University of Witwatersrand
Wits School of Public Health
Private Bag 3
Wits 2050
Johannesburg, South Africa

11 Population Council
One Dag Hammarskjold Plaza
New York, NY 10017 USA

* Members listed at the end of the manuscript

Corresponding Author: Ann Weber, PhD[1]

Email address: annweber@stanford.edu

Funding

This Series was funded by the Gender Equality, Integrated Delivery, HIV, Nutrition, Family Planning and Water Sanitation and Hygiene Program Strategy Teams at the Bill & Melinda Gates Foundation, and the United Arab Emirates Ministry of Foreign Affairs and International Cooperation, through grants to Stanford University. The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data and had final responsibility to submit the paper for publication.

Disclaimer

The views and opinions expressed in this paper are those of the authors and do not reflect the official position of any of the organizations for which the authors work.

Abstract

Despite global commitments to achieving gender equality and improving health and well-being for all, quantitative data and methods to precisely estimate the effect of gender norms on health inequities are under-developed. Nonetheless, existing global, national, and sub-national data provide key opportunities for testing associations between gender norms and health. Using innovative approaches to analysing proxies for gender norms, we generated evidence that gender norms impact the health of women and men across life stages, health sectors, and world regions. Six case studies demonstrated that: 1) gender norms are complex and may intersect with other social factors to impact health over the life course; 2) early gender-normative influences by parents and peers may have multiple and differing health consequences for girls and boys; 3) non-conformity with, and transgression of, gender norms may be harmful to health, in particular when they trigger negative sanctions; and 4) the impact of gender norms on health can be context-specific, demanding care when designing effective gender-transformative health policies and programs. Limitations of survey-based data are described that resulted in missed opportunities for exploring certain populations and domains. Recommendations for optimising and advancing research on the health impacts of gender norms are made.

Key Messages

1. Existing survey-based data can be harnessed to generate new evidence of the pervasive influence of gender norms on the health and well-being of girls, boys, women, and men across a range of health-related outcomes and the life course in high, middle, and low-income countries. While these data may be inadequate for making causal claims of the impact of specific gender norms on health, the data were sufficient to expose important gendered pathways to health and well-being. Additional opportunities remain to build on this evidence and generate new hypotheses with survey-based data.
2. By applying diverse analytical methods to different types of proxy measures for gender norms, we demonstrated that:
    a. Gender norms are complex and may intersect with other social factors to impact health over the life course;
    b. Gender-normative influences by parents and peers start early, and may have multiple short- and long-term health consequences that differ for girls and boys;
    c. Non-conformity and transgression of gender norms can be harmful to health, in particular when they trigger negative sanctions; and
    d. Gender norms are often context-specific, demanding a deeper understanding to design effective gender-transformative policies and programmes.
3. Existing survey-based data can introduce or perpetuate bias when used for studying the impact of gender norms on health:
    a. Reliance on sex-disaggregated data can result in misclassification of gender and ignores trans-gender and non-binary experiences.
    b. Datasets include rich gender-related attitude data or health-related data, but rarely both;
    c. Data are limited or non-existent for who enforces norms, how they are enforced, or what sanctions transgressors of norms may face.

d. Global datasets are generally not powered to study how gender norms intersect with strata of other social determinants of health (e.g., wealth, religion, and ethnicity) and may be missing data for entire demographic groups (e.g., boys and men, children 6-14 years, women over 49 years, gender minorities) or world regions.

e. Questions are often unbalanced by sex of the respondent (e.g., only women are asked about child health and care) and phrasing of questions frequently revealed underlying gender biases in research.

4. Future development of quantitative proxy measures for gender norms would benefit from mixed methods that utilise qualitative research to unpack the origins, preservation, and shifts in gender norms and their links with health outcomes.

5. Going forward, data on all facets of gender, including data for gender minorities, are necessary in future surveys with the above limitations addressed. To achieve these goals, collaborations are needed at multiple levels:

    a. Across disciplines to provide a conceptual bridge for effective use of data that aligns around an evidence-based research agenda;

    b. Between domain experts and gender scholars, survey designers and analysts, and community partners and policy makers to generate data systems that will enable studying health at the intersection of gender and other social determinants; and

    c. Across global data collection organisations to set standards for measuring gender, gender norms, and key demographic characteristics.

Introduction

Gender equality is a foundational human right, reflected in Sustainable Development Goal (SDG) 5, and a necessary means to achieve other SDGs, including 3, to "ensure healthy lives and promote well-being for all."[1,2] Mixed-methods studies document the consequences of gender inequality for women's and men's health.[3–6] However, quantitative data and methods are under-developed to precisely estimate these consequences and study how gender norms may contribute to health inequities. Nonetheless, existing survey-based data can be leveraged to gain important insights into pathways from gender norms to health.

Gender norms are society's spoken and unspoken rules about acceptable ways of being ~~what it means to be, or be seen as,~~ a girl or a boy, a woman or a man – how they should behave, look, and even think or feel. Gender norms are perpetuated and challenged in families, communities, schools, workplaces, institutions and the media.[3,5,7–9] These expectations start early and powerfully shape individuals' attitudes, opportunities, experiences, and behaviours, with important health consequences throughout the life course.[10]

Quantifying the effect of gender inequalities on health is challenging, partly because differences related to sex- (e.g., biological factors, including chromosomal, hormonal, and biomechanical) and gender (e.g., culturally-defined constructs associated with being female or male) are intertwined.[11–14] Globally, women outlive men by 2-4 years on average, but girls and women have a higher burden of some disabilities and morbidities.[2,15–18] These differences cannot be explained by sex alone, which we demonstrate with the 2016 Global Burden of Disease data,[19] extending work by Snow (2008).[20] We identified 15 causes of disability-adjusted life years (DALYs) that most disproportionately affected females (Figure 1a) or males (Figure 1b) globally. The >40:1 female-to-male DALY ratio from breast cancer is primarily sex-driven, whereas the ~3:1 female-to-male DALY ratio from eating disorders reflects gender-related factors.[3] Higher road traffic injuries among males, explaining nearly 4% of their all-cause age-standardised DALYs, also reflects male gender norms pertaining to driving, risk-taking, and alcohol use.[21] Sex/gender

also intersect with other social factors to impact DALY ratios. For example, given differential exposures within gendered occupations,[10] women are more vulnerable to Ebola (from nursing) in low Socio-Demographic Index (SDI) countries and men to pneumoconiosis (from mining) in high-SDI countries.[11,22]

From over a dozen case studies involving secondary analyses of existing global, national, and sub-national datasets, we selected six to present here (Table 1) based on conceptual and practical considerations (see Appendix 8 for the selection process). Conceptually, we aimed to study a range of gendered pathways to health for which evidence exists, as framed by Heise, Greene et al.[10] Our analyses were informed by feminist sociological theories of how gender norms contribute to shaping an unequal gender system that can be harmful to both women, men, boys and girls. [13,23–25] We sought to include pathways across the life course, around the world, and for diverse mental and physical health-related outcomes, despite challenges in data quality and operationalising gender norms. Following the case studies, we reflect on data opportunities and limitations, concluding with recommendations for optimising research on health impacts of gender norms.

Gendered pathways to health

We rely on sex-disaggregated data, recognising that sex and gender typically are conflated in surveys.[26,27] Additionally, existing survey data do not systematically measure gender norms, so we created proxies by aggregating individual-level data to the level of influential social or reference groups (e.g. peers). With the exception of studies 2 and 3, we aggregated gendered behaviours (what women/girls and men/boys do) or attitudes (what people believe women or men should do) to the level of a community, community cluster, or school. We then tested different pathways between gender norms and health. When data allowed, we tested how gender interacted with other analytical categories (e.g. wealth or religion) in shaping health-related social disadvantages. In case studies 1 and 5, we contrasted aggregated behaviours or attitudes for males and females to ask: "what can these differences tell us about gender norms and their implications for health?" In case studies 5 and 6, we asked of between-group variation: "can we detect

differences in individual health by the strength of the gender-normative environment?" In case studies 4 and 6, we contrasted individual behaviour with that of groups to ask: "can non-conformity with, or transgression of, the norm impact individual health—for example, can it result in harm?" Finally, in case 5, we contrasted group-level attitudes (what people should do) with the corresponding behaviour (what people actually do) to ask: "can the discordance between them impact individual health?" Only in case studies 2 and 3 do we use individual-level data for the norm, taking advantage of the normative questions: "what do you think others think about you?" to explore gender differences and ask: "can a person's belief in what others think of them affect their health?"

For each case study presented below, we link the case to a gendered pathway, including key literature; describe the data, gender norm proxy measure, and analytic approach; and present key results and insights. The case studies are arranged by life stage, from childhood, to adolescence, to early adulthood.

Case study 1. Care-seeking for childhood illness in Ethiopia

Restrictive gender norms can affect young children's health. For example, when girls are seen as a lesser financial asset than boys, parents might invest less in girls' health and education,[28–31] reflected in differences in access to care for common childhood illnesses.[32] We used geospatial information available in the Demographic and Health Survey (DHS) for Ethiopia in 2011 to examine differences in care-seeking for girls and boys <5 years (n=3,161 children in 544 villages), which we hypothesised varied within country by geographic and sociodemographic contexts.[33,34] Care-seeking was defined as medical care sought from a certified medical practitioner for symptoms of pneumonia, fever, or diarrhoea (available disease indicators) in the previous two weeks.

We aggregated individual care-seeking behaviour using geospatial hierarchical cluster analysis [35] identifying spatially proximal clusters of communities with significantly higher (hot spots) and lower (cold spots) care-seeking than the national average, separately for girls, boys, and the differential (boys

minus girls) (Appendix 1). We created a gender norms proxy of gender preference in care-seeking by assigning a yes/no indicator to communities in hot spots for differential care-seeking. We tested whether key community-level characteristics (e.g., socio-economic status, dominant religion, and vaccination rates) predicted this proxy measure.

Hot and cold spots were mapped separately for girls and boys (Figure 2). Sex-specific maps were overlaid with spatial distributions of increasingly wealthy (panels 2a and 2b) and Muslim (panels 2c and 2d) households in communities (see Appendix 1 for factor selection). Clusters of hot (or cold) spots for girls and hot (or cold) spots for boys appear in the same geographic areas, except for a cluster of hot spots for boys in the east, for which there is no equivalent for girls and where communities appear wealthier and majority Muslim. In adjusted logistic regressions of sex-specific hot spots, we found that majority Muslim (>50% of households) communities were associated with increased odds of ~~communities~~ being care-seeking hot spots for boys but decreased odds for girls ~~(Appendix Table A1.3)~~compared to communities with <50% Muslim households (Appendix Table A1.4). Differential care-seeking hot spots favouring boys ~~had a marginally significant association~~was associated ~~with mostly wealthy (>50% of households) communities, but the association was not statistically significant (OR=2·56, 95% CI 0·92, 7·12; p-value 0·071). On the other hand, differential care-seeking hot spots and~~ had a very large and significant association with majority Muslim compared to minority Muslim communities (OR=18·2~~21·49~~, 95% CI 8·72~~5~~, ~~52~~40·7~~8~~; p-value<0·0001) (Appendix Table A1.~~3~~4). ~~Communities with good vaccine coverage were also significantly associated with differential care-seeking in preference of boys (OR=2·15, 95% CS 1·17, 3·98; p-value 0·014).~~ Differential care-seeking favouring boys was also associated with mostly wealthy (>50% of households) communities, but the association was weaker and not statistically significant (OR=2·67, 95% CI 0·95, 7·46; p-value 0·062). We found no clear evidence for interaction between wealth and religion on care-seeking hot spots.

These findings suggest that, unlike reports from elsewhere,[36] poverty did not drive lower care-seeking for girls in Ethiopia. Our findings, however, are consistent with reports of son preference in other

contexts.,[37,38] ~~although the association with higher wealth was only marginally significant.~~ Notably, preferential care-seeking for boys in Ethiopia was very strongly associated with Muslim majority communities. Evidence of care seeking in favour of boys in geographically focused ~~Muslim majority~~ communities, regardless of socioeconomic status, suggests that equal access to care is insufficient in achieving gender equality and highlights the importance of local contextual variation when addressing gender norms in programming and policy.

Case studies 2 and 3. Adolescent weight control and mental health in South Africa and Brazil

Gender norms learned in the family[7,39–41] are later reinforced or challenged in the community, at school, and by the media.[9,10] Evidence suggests that internalisation of gender norms and their influence on health-related behaviours might be especially powerful during adolescence,[7–9,41–43] when important biological and psychological changes occur and many health-related behaviours are adopted.[44,45] We examine pathways through which normative pressures from parents and peers may contribute to adolescents' gendered health behaviours and differential health outcomes. We present two complementary studies together as they offered unique data on individuals' perceptions of norms around body image.

Case 2:

Known manifestations of weight concerns—for example, eating disorders—are highly gendered globally, primarily affecting girls.[3,46,47] We used prospective cohort data from South Africa (Birth-to-20)[48] to examine how early normative pressures from peers affected adolescents' later weight control behaviour, and how this association differed by sex/gender and social context. The data are from mostly Black children (N=3,273) born in Soweto-Johannesburg in the early 1990s, during a period of rapid urbanisation[48] and simultaneous emergence of eating disorders among Black girls.[49]

The gender norms measure was adolescent boys' or girls' perceptions of peers' approval of their appearance (measured on a scale of 0-never to 4-always). Adjusted linear regression models used sex-

disaggregated data from ages 13, 17, and 22 years[48] to test associations between perception and eating disorders risk (measured by the Eating Attitudes Test with three subscales: dieting, bulimia, and oral control, where higher scores mean higher risk).[50] Body satisfaction score (regarding one's own weight and appearance, where a higher score means higher satisfaction) was an intermediary factor (Table 1 and Appendix 2).

Among girls, increased perceived peer approval of their appearance between ages 13 and 17 was associated with increased body satisfaction, controlling for change in body mass index (BMI) over the same period ~~(β=3·095, 95% CI 2·199, 3·990; p-value<0·0001)~~ (β=2·567, 95% CI 1·405, 3·729; p-value<0·0001). An increase in body satisfaction, in turn, was associated with decreased dieting risk score by age 22 ~~(β= -0·061, 95% CI -0·096, -0·025; p-value=0·001)~~ (β=-0·048, 95% CI -0·088, -0·008; p-value=0·019) (Appendix Table A2.3~~1~~). This translated into a statistically significant indirect association between perceived peer approval and dieting ~~(β= -0·171, 95% CI -0·054, -0·286)~~ (β=-0·124, 95% CI -0·008, -0·240, p-value= 0.036), with similar trends for bulimia and attempts to control eating as measured by oral control scores (Appendix Figure A2.~~3~~1), and across levels of household wealth. The direct association between perceived approval and eating disorder risk was small and not statistically significant.

Boys' body satisfaction was also influenced by perceived peer opinion, but overall risk of eating disorders was not consistently influenced, with wealth having a ~~statistically significant~~ moderating role (Appendix Figure A2.~~4~~2). For boys in lower-wealth households, increased perception of peers' approval over time was associated with a ~~marginally significant~~ reduction in dieting scores, with a marked reversal of this ~~trend~~ association in higher-wealth households.

These results demonstrate the importance of peer-mediated body dissatisfaction in dieting behaviours in girls, and intersectionality of normative expectations with wealth in boys, perhaps reflecting broader media influences in wealthier households. Findings suggest that interventions aiming to reduce

adolescents' harmful weight control behaviour should engage peer networks in challenging unhealthy norms of body appearance.

Case 3:

What children believe to be their p~~P~~arents' judgments of their weight, communicated through either words ~~and~~ or actions (e.g. weight-based teasing ~~and encouragement to control weight~~)~~,~~ ~~are~~ is associated with body dissatisfaction,[51] and ~~have~~ has in turn been linked to adverse mental health outcomes. We examine the influence of normative pressure from parents in Brazil, where urban culture places high value on body appearance and is accepting of weight control behaviours.[52]

The Brazil data are from a birth cohort (N=5,249) from the city of Pelotas in 1993.[53] Here, we test the role of perceived parents' opinion of adolescent boys' and girls' weight at age 11 ('thin,' 'normal,' or 'fat') as a moderator of the effect of body dissatisfaction at age 15 (feeling fatter or thinner than ideal) on mental health at age 18. Mental health was measured using the Self-Reporting Questionnaire (SRQ) screening instrument (higher score indicates worse mental health).[54] We restricted the analytic sample to girls (n=~~1419~~1309) and boys (n=~~1245~~1113) with normal BMI at age 11 so that our gender norms proxy – perceived parental opinion for boys or girls – was unlikely to reflect genuine parental health concerns about overweight or underweight status (Appendix 3).

We found that a higher percentage of normal-BMI girls than boys reported that their parents thought they were fat at age 11 (7.1% vs 5.~~8~~6%), whereas more boys than girls reported that their parents thought they were thin (42.~~4~~6% vs 36.9~~7~~%). In sex-disaggregated regression, there was ~~a marginally significant~~some evidence for an interaction between perceived parent's opinion about weight at age 11 and body dissatisfaction at age 15 ~~(p-value =0.052)~~. Girls who thought they were fatter than ideal at age 15 had significantly poorer mental health at age 18 compared to those who were satisfied with their bodies, but only if, at age 11, they had reported that their parents thought they were fat (β=3·081, 95% CI 1·049,

5·114; p-value=0·003). In contrast, for girls who believed their parents thought they were normal or thin at age 11, feeling fatter than ideal at age 15 was not ~~significantly~~ associated with SRQ scores (Figure A3.1). We did not observe a similar pattern among boys, suggesting that parents' opinions about body image operate differently for girls' and boys' mental health. Thus, perceived parental opinion about weight appears to be a determining factor in whether girls desiring thinness impacts their mental health.

The long-term contribution of normative parental influences to girls' later mental health in Brazil suggests a more powerful influence than previously documented. These findings further emphasise the importance of multi-level interventions across influential groups, such as parents and teachers, to temper socially-driven health inequities.

Case study 4. School peer influences on adolescent health in the USA

Pressure to conform to restrictive gender norms can have profound effects on adolescents' mental health.[55–57] Negative social sanctions for transgressing norms are particularly salient during adolescence, when adolescents seek identity through group membership.[9,58] Sanctions can include bullying or ostracism by peers, and scolding or punishment by caretakers and/or teachers.[7] Here, we examine a pathway to risky health behaviours and poor outcomes from non-conformity with gender norms in schools.

We use data from the U.S. National Longitudinal Study of Adolescent to Adult Health (Add Health),[59] a nationally representative sample of adolescents aged 11-18 years (1994-1995) (n=20,745), randomly selected from 80 paired middle and high schools. The dataset lacks gender-specific attitude questions, but is rich in behavioural and health-related data. Following the work of Fleming et.al.,[60] we created a gender normativity measure for each student using a set of factors found to discriminate between binary sex assignment in the survey (Appendix Table A4.1). For the gender norms proxy, sex-specific individual scores were aggregated to the median of same-sex school-level peers. We tested non-conformity to

dominant gender norms, expressed as the difference between an individual's estimated gender normativity and the median of their same-sex school peers, on health.

For each outcome, we conducted sex-stratified piecewise linear regressions to estimate separate effects of more typically feminine and more typically masculine behaviours compared to the median of their school, controlling for an individual's own gender normativity, birth year, race/ethnicity, and school fixed effects (Appendix Table A4.6). Standardised regression coefficients are plotted for girls (Figure 3 panel a) and boys (Figure 3 panel b) (also in Appendix Table A4.6).

Multiple health-related outcomes were associated with gender norm non-conformity. Boys and girls reporting more typically 'masculine' behaviours than their same-sex peers were significantly more likely to report risky behaviours, for example engaging in delinquent behaviour (β=0·158, 95% CI 0·015, 10·531; p-value <0·0001 for girls and β=0·399, 95% CI 0·028, 14·426; p-value <0·0001 for boys). On the other hand, boys and girls reporting more typically 'feminine' behaviours, were more likely to report weight loss behaviours (β=0·228, 95% CI 0·025, 9·265; p-value <0·0001 for girls and β=0·143, 95% CI 0·018, 7·774; p-value <0·0001 for boys). Girls were more likely to report increased depressive symptoms, and suicidal ideation and attempts with increasing difference in either direction (more typically 'masculine' or 'feminine') from peers' median gender normativity score. Results were similar controlling for household socioeconomic status (Appendix Table A4.7).

In summary, US students at the extremes of a gender-normative measure relative to other students in their school may suffer multiple health-related effects. Negative sanctions from gender-norm dominant peers may be one of the paths through which these associations operate. These results highlight the need to address stigma and negative behavioural and mental health consequences associated with gender non-conformity in schools.

Case study 5. Premarital sex and HIV status in Zambia

Sub-Saharan Africa has the highest prevalence of human immunodeficiency virus (HIV) infection globally, with new cases concentrated among adolescents[44] and disproportionately among girls.[31,61] Gender norms and power imbalances play a key role in HIV acquisition,[62–64] as they impact, for instance, condom access and use.[62,63] In the USA, embarrassment may prevent adolescents from receiving HIV information, seeking contraception, using condoms, or accessing care.[65,66]

We examine a gendered pathway to HIV infection among youth in Zambia through community expectations of appropriate sexual behaviour.[67,68] Where social norms against premarital sex exist, we hypothesised that youth engaging in premarital sex would refrain from talking about it (with peers, parents, or health professionals), reducing their ability to learn about and access HIV protection and increasing their acquisition risk. We also hypothesised a greater impact on girls than boys, partly because of double standards[10,69] regarding appropriate sexual behaviour.

We analysed data for young women (n=1669) and men (n=1285) (ages 15-24 years) from the 2007 DHS in Zambia, one of six countries with HIV status information and balanced questions about expectations around premarital sex (Appendix 5). The gender norms proxy was adult (ages 25-49) women and men's attitudes about premarital sex, obtained by aggregating sex-specific data to 18 regional and urban-rural strata. We tested the effect of adult non-compliance with norms for premarital sex, expressed as the discordance between adult attitudes and their behaviours (believing premarital sex to be wrong, but engaging in it), on HIV acquisition risk among youth (n=2954).

Attitudes towards premarital sex did not vary substantially by sex or region in Zambia and were conservative: more than 80% of adults disapproved of premarital sex in most regions (Figure 4, panel a). In contrast, attitudes and behaviours were mostly discordant for men (most disapproved of premarital sex, but were assessed as having engaged in it, panel b), whereas women were more likely to be concordant

(most disapproved of premarital sex and refrained from it). Women's perceptions of what most other women did (descriptive norms of high perceived prevalence of premarital sex) were discordant with their own behaviours (lower prevalence of premarital sex, panel c). Panel d illustrates substantial heterogeneity in HIV prevalence among youth (15-24 years) across Zambia (range 3-27%), disproportionately affecting young women in urban regions.

At the regional level, an increasing proportion of adult women (25-49 years) who refrained from engaging in premarital sex was associated with reduced HIV prevalence among adolescent women (Pearson correlation, rho=-0·43; p-value=0·077), while conservative attitudes were not. Importantly, discordance among adult women was strongly correlated with adolescent women's HIV prevalence (rho=0·63; p-value=0·005), explaining an additional 20% of the variation in adolescent women's HIV status over behaviour alone. Furthermore, in sex-stratified Poisson regressions, we found that a 10% increase in discordance among adult women or adult men was associated with a ~~2.4~~27% (RR=1·~~02~~427, 95% CI 1·~~01011~~,; 1·~~03845~~; p-value=0·001) or ~~2.5~~28% (RR=1·~~025~~28; 1·~~005~~05, 1·~~046~~56; p-value=0·015) increase, respectively, in individual-level relative risk of HIV for adolescent women, controlling for demographic and regional-level factors (Appendix 5). Risks were similar for adolescent men, but not statistically significant.

These results illustrate that gender norm non-compliance can harm health, here the risk of HIV infection, with potentially fatal consequences. Given sexual double standards,[10,69] young women may especially avoid seeking information, negotiating condom use, or seeking care to minimise risks of premarital sex, as they may anticipate heightened disapproval, relative to men. Efforts to protect women from harm associated with sexual activity should consider the normative environment in which adolescents' sexual relationships take place.

Case study 6. Women working outside the home and intimate partner violence in Nigeria

Gender norms intersect with power as adolescents move into early adulthood,[5,7,8,43,70] with unequal power relations shaping and being shaped by gender inequalities and restrictive gender norms.[10,13] Those in power benefit from, and seek to uphold, the existing social order by (consciously or unconsciously) sanctioning non-compliers.[71,72] We examine a pathway through which gendered power disparities can generate punishment (privately, at home) for women who violate the gender order by working outside the home.

Evidence is mixed on whether female labour force participation (FLFP) increases[73–76] or reduces[77,78] women's risk of intimate partner violence (IPV) in low gender-equality contexts, as IPV largely takes place in private. FLFP can be protective for working women in countries where most women work, but may be a risk factor for IPV in countries where most women do not.[78,79] We tested whether women who work outside the home are at increased IPV risk relative to women who do not in two types of communities in Nigeria: communities where few women work outside the home and communities where FLFP is more normative.

We used data from the 2014 cluster sample design Violence against Children Survey (VACS) on experience of IPV for female youth (n=1,633, ages 13-24) (Appendix 6). FLFP was based on self-reported work outside the home in the last week. We used intraclass correlation coefficients (ICC) to detect that FLFP was clustered at the community level for girls (but not boys), with sufficient heterogeneity across communities to test our hypothesis. Assuming equal economic opportunities for work across communities, a low proportion of young women engaging in work outside the home was our gender norms proxy reflecting restrictive norms around women's mobility and opportunities to earn income. Communities were then classified as either: 1) FLFP-high (assumed absence of restrictive norms around FLFP), or 2) FLFP-low (assumed presence of norms sanctioning FLFP), based on a data-driven

cut-point of 28% of female respondents engaging in outside labour. Results were robust to different cut-points (data not shown).

There were no statistically significant differences in overall past-year exposure to sexual or physical IPV for all women between the two community types (adjusted Wald tests [FLFP-high 7·3% (1·16); FLFP-low 7·9% (1·50); p-value=0·733]). Using logistic regression controlled for age, marital status, and having ever attended school, we found that women who worked in FLFP-low communities had significantly higher odds of experiencing past-year IPV compared to non-working women [OR=2·381, 95% CI 1·292, 4·389; p-value=0·006]. However, in FLFP-high communities, women's IPV risk did not differ by working status (Appendix Table A6.4).

The increased risk of IPV exposure for working women in FLFP-low communities suggests that some male partners may use IPV to punish women for transgressing gender norms around work and the perceived threat to their masculine role as breadwinner or power-holder. Although early transgressors of restrictive norms may experience IPV as a consequence, they may also initiate long-term norm changes in ways that improve employment opportunities and health for future generations.[80] We examine elsewhere the implications of gender norms for FLFP and women's health across geo-cultural contexts[81] and time.[82]

These findings have important implications for ~~programs~~ interventions ~~navigating~~ at the intersection of gender equality and global health and development—for example, efforts to empower women through employment or micro-finance of small businesses. When instituting such empowerment programmes, risks of harm to those encouraged to challenge restrictive gender norms must be anticipated, and harm prevention and mitigation strategies implemented for effective reduction in gender inequalities and health inequities.

Opportunities and challenges

Our case studies provided practical opportunities to conduct gender norm-health research using existing survey data in new ways. For example, geospatial clustering in case 1 revealed regional variation in gender norms where sex intersected with religious identity to produce large inequities in healthcare seeking – a finding that individual-level analyses might miss. Clustering communities together overcame the challenge of small numbers (i.e., precision) when estimating group-level behaviours for communities with few sick children. This innovative approach to identifying gender inequities could be extended to other health-related indicators and countries.

The inclusion of a targeted question in case studies 2 and 3 about 'what adolescents thought that others thought' was useful for estimating the normative influence of peers and parents. Similarly targeted questions could be added with limited additional expense to future surveys. In case 4, the construction of a gender normativity index enabled the use of a dataset rich in measures of gender-related behaviours to study gender non-conformity and health. This novel approach could be generalised to datasets such as the Global school-based student health survey to expand this exploration in diverse contexts.

The measure of discordance between group-level attitudes and behaviours related to premarital sex in case 5 disrupted the common practice of using only attitudes or only behaviours as gender norm proxies. Contrasting other matched attitude-behaviour pairs in this way could generate additional new insights for gendered pathways to health, as shown here for the acquisition of HIV. Finally, case 6 demonstrates how ICC, which is traditionally used to estimate effective sample size in clustered study designs, can be reinterpreted to identify sufficient clustering of behaviours to study within-country variation in gender norms.

Nevertheless, we encountered multiple data limitations, not the least of which was relying on sex-disaggregated data to study gender. In recent decades, global health leaders have increasingly recommended incorporation of gender in data systems.[12,83–89] A comprehensive United Nations report on

gender statistics recommended that data should systematically be sex-stratified; measure gender facets, including norms and relations; reflect the diversity of women and men, capturing multi-dimensional aspects of their lives; and be free of gender stereotypes and biases.[88] While these guidelines provide a useful framework for collecting gender-sensitive data, none of the 17 publicly available data sources we explored (Appendix Table A8.1~~7~~) were designed accordingly. The substitution of a binary sex indicator for gender in sex-disaggregated data represents a missed opportunity to study gender and health along a continuum of experiences and may have introduced important misclassification biases in our analyses.

Moreover, many datasets lacked the combination of gender-related attitudes or behaviours and health outcomes required for understanding pathways between them. Even when both were available, data were often missing for certain demographic groups or regions of the world. For example, DHS represent low- and middle-income countries and data were often missing for men (e.g., questions on child care), women (e.g., questions on some sexual practices), or certain age groups (e.g., children 6-14 years and women over 49), which can bias data interpretation. In some cases, the available proxy was perhaps too distal from the health outcome of interest, or confounded by intermediate factors, to detect an association (e.g., between attitudes around IPV and childhood malnutrition).[90]

Additional data limitations included the inability to stratify samples by subgroups, both because of lack of indicators (e.g., missing race/ethnicity information) and small samples. Attempts to disaggregate national survey data to sub-national levels or across socio-economic strata decreased statistical power, limiting our capacity to study impacts of intersecting disadvantage with precision.

Notably, we encountered survey questions that belied gender-biased assumptions in their construction. For example, we used the rich attitudinal data in the World Values Survey (WVS) to explore adult self-rated health and gender norms around employment. However, the employment status question cannot account for cross-cultural differences in the meaning of self-employment, and includes the gender-biased term “housewife” as one of its English-version response categories. Forty-three of 46 surveys back-

translated to English used a housewife-like phrase or word (21 of 24 languages and 33 of 36 countries) as opposed to a gender-neutral description (Appendix Table A7.3~~7~~). Such variation made the category unreliable for cross-national comparisons and likely biased. Additionally, phrasing of attitudinal questions, such as "Pre-school children suffer with a working mother," communicates the stereotype that mother's role is at home as caregiver while father's employment-related absence is inconsequential for young children. It is also unclear whether the question refers to a situation where both parents work, or only the mother versus the father works. Furthermore, questions phrased with the terms "wife" or "husband" suggest that the questions only apply to married couples in heterosexual unions.

Finally, women and men may answer survey questions based on gendered expectations of what they think they should say rather than on their lived experiences, particularly around such gender-charged topics as sexual behaviour or eating disorders. Potentially biased responses may have led us to reproduce current, potentially biased understandings of gendered behaviour and health risk, while missing important at-risk groups.

Combined, these data limitations hindered our exploration of how, and by whom, norms are enforced and the differential impacts of norm violations across the life course and world regions. Heise, Greene et al. ~~demonstrated~~ argued that gender "biases can be manifested and reinforced by research methodologies".[10] ~~In this paper, w~~While publicly available survey data provided many opportunities for testing hypotheses about gender norms and health, care is required to avoid introducing or perpetuating bias when constructing and using gender norms proxies from these data.

Research agenda

In future research, we join many others in advocating for collecting survey-based data on all facets of gender, including data for gender minorities.[12,83–89] We also advocate for balanced survey data in which men and women are equally represented across age groups and asked the same unbiased attitudinal and

behavioural questions, enabling gender-comparative research. Given constrained resources, we recognise that choices must be made in designing surveys, but each confers trade-offs that should be analysed from an intersectional lens encompassing gender. If certain domains are assumed unimportant (e.g., childcare provided by men) and hence not measured, then we will not be able to assess or effect change.[91] Data that reflect society not only as it is, but also as we aspire for it to be, are critical for monitoring progress on SDGs. Identifying and better measuring current and evolving gender norms across cultures, life stages, and areas of society will enable more robust study of gender norms and health.

In addition to more gender-sensitive data, we require more research on gendered pathways to health, including integrating qualitative research to unpack the origins, preservation, and shifts in gender norms. The collection of harmonised and consistent data across contexts and over time (e.g., standards for measuring gender and gender norms across global surveys), combined with longitudinal methods, would allow for cross-national comparisons, assessments of cohort effects and causal impact, and monitoring of gender norm evolution. Methods that overlay different types of data, such as survey-based and geospatial data, could utilise external factors (e.g., climate change and economic shocks) to identify locations of gender-based discrimination. Machine learning algorithms and natural language processing could offer novel approaches to eliminating gender-related biases coded in large existing datasets.

Finally, we advocate for enhanced collaborations across the humanities and social and health sciences to provide conceptual bridges for effective data use around an evidence-based research agenda. Representation from domain experts and gender scholars, survey designers and analysts, and community partners and policy makers will allow for data systems that enable studying health at the intersection of gender and other social determinants (e.g., race, religion, and social class). Identifying mechanisms for safely sharing and analysing survey datasets is critical for safeguarding privacy while enabling new opportunities to study this intersectionality in global health research.

Conclusion

A variety of analytic tools applied to existing survey-based data across six case studies examined how restrictive gender norms can harm the health of women and men, boys and girls, across diverse settings and outcomes. We demonstrated how to construct creative gender norm proxies and conduct analyses using a variety of methods to gain novel insights into links between gender norms and health using available survey data. We also presented key limitations to advancing the field.

Four key findings emerged that have important implications for programmatic practice and policy. First, as ~~in~~ the case study on care-seeking for childhood illness in Ethiopia shows, gender norms may intersect with other social determinants to impact health, sometimes in unexpected ways, deviating from what practitioners and policy-makers might intuitively anticipate. Second, as evidence in Brazil and South Africa suggests, early gender normative influences may affect health in different ways for boys and girls, and differentially by family context. Third, as the Add Health data in the US and the VACS data in Nigeria highlight, gender non-conformity and norm transgression may be harmful to health, particularly when challenging power relations and triggering negative sanctions. Finally, as shown with proxy measures across case studies, the impact of gender norms can be highly context-specific. Therefore, generalisations around gender norms can be counterproductive, misleading, or even harmful. Ecological studies (e.g., with national indicators of gender inequality), while informative for hypothesis generation, belie the complexity and importance of local factors that influence relationships between gender norms and health. A deep understanding of sociocultural contexts, aided by qualitative research, is required to design effective prevention and mitigation strategies for socially-driven health inequities, and ongoing monitoring must be in place to identify, support and protect those who challenge restrictive gender norms and existing gender-based power differentials. Public health programs and policies that are locally relevant while globally active are central to achieving both gender equality and health. Progress can be accelerated through improved qualitative and quantitative data collection, analysis, and interpretation that accounts for the pervasive role of gender norms in shaping human health and well-being.

563

564 **Author Contributions**

565 AW, BC and VM worked closely with analysts and data owners to conceive, plan, and interpret results
566 from the case studies. They framed, drafted, and revised the manuscript.

567 GLD was the Principal Investigator of *The Lancet* Series on Gender and Health project and implemented
568 the multiple data contributor/partnership for the case studies. He also worked closely with the analysts
569 and data owners to conceive, plan, and interpret results from the case studies, as well as providing critical
570 input on framing, review and edits to the manuscript.

571 The following authors worked on one or more case studies, including data analyses and writing methods
572 and results (e.g., for appendices), contributed to the interpretation of the results for case studies, as well as
573 a critical review of the manuscript: for case study 1: PL (primary) with AJDB, IMG, EH, and SA, and the
574 support of working group members in Pelotas, Brazil; case study 2: SA (primary) with LR and SAN; case
575 study 3: SA (primary) with CGV and RB; case study 4: BD (primary) with JN, HB, and SA; case study 5:
576 IMG (primary) with EH and PL; case study 6: IS and LS (primary) with EH and IMG.

577 Additional input for case studies received from TN, NH, SC, and KM, as well as review and edits to the
578 manuscript.

579 DB performed literature searches for individual case studies and the overall paper and contributed to the
580 writing and review of the manuscript.

581 RG performed the analyses for the survey question translation example, prepared the data and created the
582 word clouds for the DHS modules, and contributed to the methods documentation.

583 MC provided critical input on framing, review and edits to the manuscript.

584

585 **Declaration of Interests**

586 The authors declare they have no conflicts of interest. The views expressed are those of the authors and
587 are not necessarily those of the Bill & Melinda Gates Foundation and the United Arab Emirates. As
588 corresponding author, AMW states that she had full access to all data and final responsibility to submit
589 for publication.

590

Acknowledgments

We thank our working group members[a] for their contributions to the six case studies, as well as our collaborators who worked in parallel on case studies not included in this paper. We also thank the members of the Gender Equality, Norms and Health Steering Committee[b] for their insights and critical review throughout the development of the paper.

[a]Working Group

Kathleen M. Harris, PhD, University of North Carolina at Chapel Hill

Carolina Coll, PhD, Federal University of Pelotas, Brazil

Janaina Calu Costa, MS, Federal University of Pelotas, Brazil

Leonardo Ferreira, MS, Federal University of Pelotas, Brazil

Franciele Hellwig, Federal University of Pelotas, Brazil

Eran Bendavid, MD, Stanford University

Marcella Alsan, MD, PhD, Stanford University

~~Elias Nosrati, University of Cambridge~~

~~Larry King, PhD, University of Massachusetts, Amherst~~

Sarah Hawkes, MD, PhD, University College London

James Zou, PhD, Stanford University

Londa Schiebinger,PhD, Stanford University

Lori Heise, PhD, Johns Hopkins Bloomberg School of Public Health

[b]Gender Equality, Norms and Health Steering Committee:

Gary Darmstadt, MD (Chair); Margaret Greene, PhD; Sarah Hawkes, PhD; Lori Heise, PhD; Sarah Henry, MPH; Jody Heymann, PhD; Jeni Klugman, PhD; Ruth Levine, PhD; Anita Raj, PhD; Geeta Rao Gupta, PhD

References

1 Magar V. Gender, health and the Sustainable Development Goals. *Bulletin of the World Health Organization* 2015; **93**: 743–743.

2 UN Women. Turning promises into action: Gender equality in the 2030 Agenda for Sustainable Development. 2018 http://www.unwomen.org/en/digital-library/sdg-report (accessed Oct 1, 2018).

3 Connell R. Gender, health and theory: Conceptualizing the issue, in local and world perspective. *Social Science & Medicine* 2012; **74**: 1675–83.

4 Langer A, Meleis A, Knaul FM, *et al.* Women and health: The key for sustainable development. *The Lancet* 2015; **386**: 1165–210.

5 Munoz Boudet, Ana Maria, Patetsch, Patti, Turk, Carolyn, Thumala, Angelica. On norms and agency: Conversations about gender equality with women and men in 20 countries. 2012.

6 Sen G, Östlin P. Gender inequity in health: why it exists and how we can change it. 2008.

7 Basu S, Zuo X, Lou C, Acharya R, Lundgren R. Learning to be gendered: Gender socialization in early adolescence among urban poor in delhi, india, and shanghai, china. *Journal of Adolescent Health* 2017; **61**: S24–9.

8 Lundgren R, Burgess S, Chantelois H, Oregede S, Kerner B, Kågesten AE. Processing gender: Lived experiences of reproducing and transforming gender norms over the life course of young people in Northern Uganda. *Culture, Health & Sexuality* 2018; : 1–17.

9 Kågesten A, Gibbs S, Blum RW, *et al.* Understanding factors that shape gender attitudes in early adolescence globally: a mixed-methods systematic review. *PLOS ONE* 2016; **11**: e0157805.

10 Heise L, Greene, M, Opper N, et al. Gender inequality and restrictive gender norms: Framing the challenges to health. *Lancet* 2019.

11 Krieger N. Genders, sexes, and health: What are the connections—and why does it matter? *International Journal of Epidemiology* 2003; **32**: 652–7.

12 Schiebinger, L., Klinge, I., Sánchez de Madariaga, I., Paik, H. Y., Schraudner, M., and Stefanick, M. (Eds.) (2011-2015). Analyzing how sex and gender interact | Gendered Innovations. http://genderedinnovations.stanford.edu/methods/how.html (accessed Oct 1, 2018).

13 Sen G, Östlin P, George, A. Unequal, unfair, ineffective and inefficient: Gender inequity in health—why it exists and how we can change it. *Global Public Health* 2008; **3**: 1–12.

14 Springer KW, Mager Stellman J, Jordan-Young RM. Beyond a catalogue of differences: A theoretical frame and good practice guidelines for researching sex/gender in human health. *Social Science & Medicine* 2012; **74**: 1817–24.

15 Ahmed T, Vafaei A, Auais M, Guralnik J, Zunzunegui MV. Gender roles and physical function in older adults: Cross-sectional analysis of the International Mobility in Aging Study (IMIAS). *PLoS ONE* 2016; **11**: e0156828.

16 Cullen MR, Baiocchi M, Eggleston K, Loftus P, Fuchs V. The weaker sex? Vulnerable men and women's resilience to socio-economic disadvantage. *SSM - Population Health* 2016; **2**: 512–24.

17 Hosseinpoor AR, Stewart Williams J, Amin A, *et al.* Social determinants of self-reported health in women and men: Understanding the role of gender in population health. *PLoS ONE* 2012; **7**: e34799.

18 Mokdad AH, Forouzanfar MH, Daoud F, *et al.* Global burden of diseases, injuries, and risk factors for young people's health during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2016; **387**: 2383–401.

19 Institute for Health Metrics and Evaluation. Global health data exchange: Global Burden of Disease (GBD) Results Tool. 2018. http://ghdx.healthdata.org/gbd-results-tool (accessed Oct 1, 2018).

20 Snow RC. Sex, gender, and vulnerability. *Glob Public Health* 2008; **3 Suppl 1**: 58–74.

21 World Health Organization. Gender and road traffic injuries. 2002.

22 Messing K, Östlin P, World Health Organization. Gender equality, work and health: A review of the evidence. Geneva: World Health Organization, 2006.

23 Courtenay WH. Constructions of masculinity and their influence on men's well-being: a theory of gender and health. *Social science & medicine* 2000; **50**: 1385–401.

24 Fleming PJ, Agnew-Brune C. Current trends in the study of gender norms and health behaviors. *Current opinion in psychology* 2015; **5**: 72–7.

25 Lazar M. Feminist critical discourse analysis: Gender, power and ideology in discourse. Springer, 2005.

26 Phillips SP. Defining and measuring gender: A social determinant of health whose time has come. *International Journal for Equity in Health* 2005; **4**. DOI:10.1186/1475-9276-4-11.

27 Morgan R, George A, Ssali S, Hawkins K, Molyneux S, Theobald S. How to do (or not to do)… gender analysis in health systems research. *Health Policy and Planning* 2016; **31**: 1069–78.

28 Das Gupta M, Zhenghua J, Bohua L, Zhenming X, Chung W, Hwa-Ok B. Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *Journal of Development Studies* 2003; **40**: 153–87.

29 Diamond-Smith N, Luke N, McGarvey S. 'Too many girls, too much dowry': Son preference and daughter aversion in rural Tamil Nadu, India. *Cult Health Sex* 2008; **10**: 697–708.

30 Hesketh T. Selecting sex: the effect of preferring sons. *Early Hum Dev* 2011; **87**: 759–61.

31 Jayachandran S. The roots of gender inequality in developing countries. *Annual Review of Economics* 2015; **7**: 63–88.

32 Calu Costa J, Wehrmeister FC, Barros AJ, Victora CG. Gender bias in careseeking practices in 57 low- and middle-income countries. *J Glob Health* 2017; **7**: 010418.

682 33 Bohn JA, Kassaye BM, Record D, *et al.* Demographic and mortality analysis of hospitalized children at
683 a referral hospital in Addis Ababa, Ethiopia. *BMC Pediatr* 2016; **16**: 168.

684 34 Gelaw YA, Biks GA, Alene KA. Effect of residence on mothers' health care seeking behavior for
685 common childhood illness in Northwest Ethiopia: A community based comparative cross-sectional
686 study. *BMC Res Notes* 2014; **7**: 705.

687 35 Getis A, Ord JK. The analysis of spatial association by use of distance statistics. In: Perspectives on
688 Spatial Data Analysis. Springer, 2010: 127–45.

689 36 Sen G, Iyer A. Who gains, who loses and how: Leveraging gender and class intersections to secure
690 health entitlements. *Soc Sci Med* 2012; **74**: 1802–11.

691 37 Hopcroft RL. Parental status and differential investment in sons and daughters: Trivers-Willard
692 revisited. *Social Forces* 2005; **83**: 1111–36.

693 38 Veller C, Haig D, Nowak MA. The Trivers-Willard hypothesis: Sex ratio or investment? *Proc Biol Sci*
694 2016; **283**. DOI:10.1098/rspb.2016.0126.

695 39 Tenenbaum HR, Leaper C. Are parents' gender schemas related to their children's gender-related
696 cognitions? A meta-analysis. *Dev Psychol* 2002; **38**: 615–30.

697 40 van der Pol LD, Groeneveld MG, van Berkel SR, *et al.* Fathers' and mothers' emotion talk with their
698 girls and boys from toddlerhood to preschool age. *Emotion* 2015; **15**: 854–64.

699 41 Yu C, Zuo X, Blum RW, *et al.* Marching to a different drummer: A cross-cultural comparison of young
700 adolescents who challenge gender norms. *J Adolesc Health* 2017; **61**: S48–54.

701 42 De Meyer S, Kågesten A, Mmari K, *et al.* 'Boys should have the courage to ask a girl out': Gender
702 norms in early adolescent romantic relationships. *J Adolesc Health* 2017; **61**: S42–7.

703 43 Vu L, Pulerwitz J, Burnett-Zieman B, Banura C, Okal J, Yam E. Inequitable gender norms from early
704 adolescence to young adulthood in Uganda: Tool validation and differences across age groups. *J*
705 *Adolesc Health* 2017; **60**: S15–21.

706 44 Patton GC, Sawyer SM, Santelli JS, *et al.* Our future: A Lancet commission on adolescent health and
707 wellbeing. *Lancet* 2016; **387**: 2423–78.

708 45 Ricciardelli LA, McCabe MP, Holt KE, Finemore J. A biopsychosocial model for understanding body
709 image and body change strategies among children. *Journal of Applied Developmental Psychology*
710 2003; **24**: 475–95.

711 46 Bair A, Steele JR, Mills JS. Do these norms make me look fat? The effect of exposure to others' body
712 preferences on personal body ideals. *Body Image* 2014; **11**: 275–81.

713 47 Brady J, Iwamoto DK, Grivel M, Kaya A, Clinton L. A systematic review of the salient role of feminine
714 norms on substance use among women. *Addict Behav* 2016; **62**: 83–90.

715 48 Richter L, Norris S, Pettifor J, Yach D, Cameron N. Cohort profile: Mandela's children: The 1990 birth
716 to twenty study in South Africa. *International Journal of Epidemiology* 2007; **36**: 504–11.

717 49 Szabo CP, Berk M, Tlou E, Allwood CW. Eating disorders in black South African females. *South African*
718 *Medical Journal* 1995; **85**. https://www.ajol.info/index.php/samj/article/view/156125 (accessed Oct
719 1, 2018).

720 50 Garner DM, Olmsted MP, Bohr Y, Garfinkel PE. The eating attitudes test: Psychometric features and
721 clinical correlates. *Psychol Med* 1982; **12**: 871–8.

722 51 Reel J, Voelker D, Greenleaf C. Weight status and body image perceptions in adolescents: current
723 perspectives. *Adolescent Health, Medicine and Therapeutics* 2015; : 149.

724 52 Goldenberg M. The body as capital: Understanding Brazilian culture. *Vibrant - Vibrant Virtual*
725 *Brazilian Anthropology* 2010; **7**: 220–38.

726 53 Victora CG, Hallal PC, Araújo CL, Menezes AM, Wells JC, Barros FC. Cohort profile: The 1993 Pelotas
727 (Brazil) birth cohort study. *Int J Epidemiol* 2008; **37**: 704–9.

728 54 Beusenberg M, Orley JH, Health WHOD of M. A user's guide to the Self-Reporting Questionnaire
729 (SRQ). Geneva: Geneva : World Health Organization, 1994
730 http://apps.who.int/iris/handle/10665/61113 (accessed Oct 1, 2018).

731 55 Gordon AR, Krieger N, Okechukwu CA, *et al.* Decrements in health-related quality of life associated
732 with gender nonconformity among U.S. adolescents and young adults. *Qual Life Res* 2017; **26**: 2129–
733 38.

734 56 Roberts AL, Rosario M, Slopen N, Calzo JP, Austin SB. Childhood gender nonconformity, bullying
735 victimization, and depressive symptoms across adolescence and early adulthood: An 11-year
736 longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry* 2013; **52**: 143–
737 52.

738 57 Valentine SE, Shipherd JC. A systematic review of social stress and mental health among transgender
739 and gender non-conforming people in the United States. *Clin Psychol Rev* 2018; published online
740 March 28. DOI:10.1016/j.cpr.2018.03.003.

741 58 van de Bongardt D, Reitz E, Sandfort T, Deković M. A meta-analysis of the relations between three
742 types of peer norms and adolescent sexual behavior. *Personality and Social Psychology Review* 2015;
743 **19**: 203–34.

744 59 Harris KM. The Add Health study: Design and accomplishments. 2013. DOI:10.17615/c6tw87.

745 60 Fleming PJ, Harris KM, Halpern CT. Description and Evaluation of a Measurement Technique for
746 Assessment of Performing Gender. *Sex Roles* 2017; **76**: 731–46.

747 61 Sia D, Onadja Y, Nandi A, Foro A, Brewer T. What lies behind gender inequalities in HIV/AIDS in sub-
748 Saharan African countries: Evidence from Kenya, Lesotho and Tanzania. *Health Policy Plan* 2014; **29**:
749 938–49.

62 Gottert A, Barrington C, Pettifor A, *et al.* Measuring men's gender norms and gender role conflict/stress in a high HIV-prevalence South African setting. *AIDS Behav* 2016; **20**: 1785–95.

63 Haberland. The case for addressing gender and power in sexuality and HIV education: A comprehensive review of evaluation studies. *International Perspectives on Sexual and Reproductive Health* 2015; **41**: 31.

64 Wingood GM, DiClemente RJ. Application of the theory of gender and power to examine HIV-related exposures, risk factors, and effective interventions for women. *Health Education & Behavior* 2000; **27**: 539–65.

65 Commendador KA. Parental influences on adolescent decision making and contraceptive use. *Pediatric Nursing* 2010; **36**: 147–56.

66 Daley AM. What influences adolescents' contraceptive decision-making? A meta-ethnography. *J Pediatr Nurs* 2014; **29**: 614–32.

67 Exavery A, Mubyazi GM, Rugemalila J, *et al.* Acceptability of condom promotion and distribution among 10-19 year-old adolescents in Mpwapwa and Mbeya rural districts, Tanzania. *BMC Public Health* 2012; **12**: 569.

68 Fearon E, Wiggins RD, Pettifor AE, Hargreaves JR. Is the sexual behaviour of young people in sub-Saharan Africa influenced by their peers? A systematic review. *Social Science & Medicine* 2015; **146**: 62–74.

69 Emmerink PMJ, Vanwesenbeeck I, van den Eijnden RJJM, Ter Bogt TFM. Psychosexual correlates of sexual double standard endorsement in adolescent sexuality. *J Sex Res* 2016; **53**: 286–97.

70 Madiba S, Ngwenya N. Cultural practices, gender inequality and inconsistent condom use increase vulnerability to HIV infection: narratives from married and cohabiting women in rural communities in Mpumalanga province, South Africa. *Glob Health Action* 2017; **10**: 1341597.

71 Cislaghi B, Heise L. Four avenues of normative influence: A research agenda for health promotion in low and mid-income countries. *Health Psychology* 2018; **37**: 562–73.

72 Cislaghi B, Heise LL. Measuring gender-related social norms: Report of a meeting. Baltimore, MD: Learning Group on Social Norms and Gender-based Violence of the London School of Hygiene and Tropical Medicine, 2016.

73 Chin Y-M. Male backlash, bargaining, or exposure reduction? Women's working status and physical spousal violence in India. *Journal of Population Economics* 2012; **25**: 175–200.

74 Heath R. Women's access to labor market opportunities, control of household resources, and domestic violence. ; : 27.

75 Hynes ME, Sterk CE, Hennink M, Patel S, DePadilla L, Yount KM. Exploring gender norms, agency and intimate partner violence among displaced Colombian women: A qualitative assessment. *Global Public Health* 2016; **11**: 17–33.

76 Koenig MA, Ahmed S, Hossain MB. Women's status and domestic violence in rural Bangladesh: Individual- and community-level effects. 2003; **40**: 20.

77 Kabeer N. Conflicts over credit: Re-evaluating the empowerment potential of loans to women in rural Bangladesh. *WORLD DEVELOPMENT*; : 22.

78 Vyas S, Watts C. How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published evidence. *Journal of International Development* 2009; **21**: 577–602.

79 Heise LL, Kotsadam A. Cross-national and multilevel correlates of partner violence: An analysis of data from population-based surveys. *The Lancet Global Health* 2015; **3**: e332–40.

80 Schuler SR, Nazneen S. Does intimate partner violence decline as women's empowerment becomes normative? Perspectives of Bangladeshi women. *World Development* 2018; **101**: 284–92.

81 Hallgren E, et al. Global assessment of the relationship between gender differences in self-reported health and regional differences in female employment rate and attitudes about female labour force participation. *(in preparation)*.

82 Falconi A, et al. Shifts in gender norms in women's paid employment during the world war ii era and later life health: An application of latent transition analysis. *(in preparation)*.

83 Buvinic M, Furst-Nichols R, Koolwal G. Mapping gender data gaps. 2014.

84 Government of Canada CI of HR. How to integrate sex and gender into research. 2018; published online Feb 12. http://www.cihr-irsc.gc.ca/e/50836.html (accessed Oct 1, 2018).

85 Lawrence K, Rieder A. Methodologic and ethical ramifications of sex and gender differences in public health research. *Gend Med* 2007; **4 Suppl B**: S96-105.

86 Nieuwenhoven L, Klinge I. Scientific excellence in applying sex- and gender-sensitive methods in biomedical and health research. *J Womens Health (Larchmt)* 2010; **19**: 313–21.

87 Nowatzki N, Grant KR. Sex is not enough: The need for gender-based analysis in health research. *Health Care Women Int* 2011; **32**: 263–77.

88 United Nations Children's Fund (UNICEF). Harnessing the power of data for girls: Taking stock and looking ahead to 2030. 2016.

89 Wizemann, Theresa M; Institute of Medicine. Sex-specific reporting of scientific research: A workshop summary. Washington, D.C.: National Academies Press, 2012 DOI:10.17226/13307.

90 Chai J, Fink G, Kaaya S, *et al.* Association between intimate partner violence and poor child growth: results from 42 demographic and health surveys. *Bulletin of the World Health Organization* 2016; **94**: 331–9.

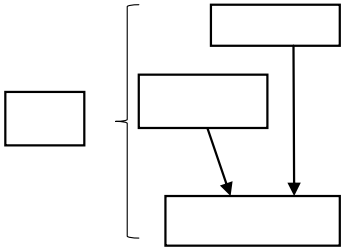
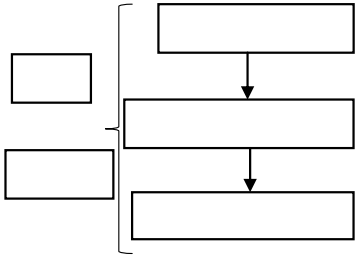
91 Hunt K, Annandale E. Just the job? Is the relationship between health and domestic and paid work gender-specific? *Sociology of Health and Illness* 1993; **15**: 632–64.

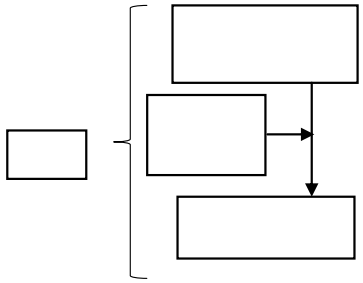
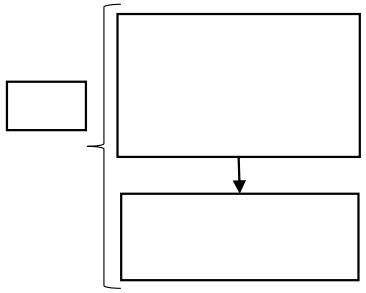
819 92 Graham AJ, Atkinson PM, Danson FM. Spatial analysis for epidemiology. *Acta Trop* 2004; **91**: 219–25.

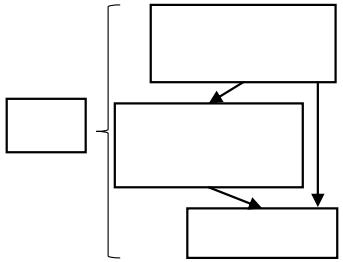
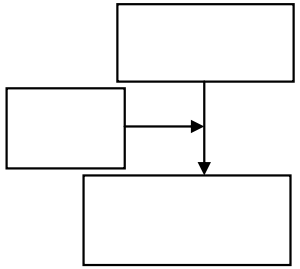
820

Tables, Figures & Panels

Table 1: Overview of case study analyses

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 1 Differential care-seeking of ill children	Gender differences in access to care	DHS, Ethiopia, 2011	Children, 0-5 y, who were ill in prior 2 weeks (n=3,161 children in 544 villages)	Indicator of communities being in a hot spot (compared to national average) for differential care-seeking for boys minus girls (proxy for gender preference)	What community factors best predict hot spots for differential care-seeking for boys vs. girls?		Differential care-seeking increased with increasing percentages of wealthy and Muslim households in communities. Differential care-seeking was greatest in communities that were both wealthy and Muslim-majority.
# 2 Community peer influence and eating disorders	Gendered health behaviours	Birth-to-20 Cohort, Soweto-Johannesburg, 1994	Male and female youth, 13-22 y (n=3273)	Individual-level perception of peers' approval of their appearance	Do adolescent perceptions of peers' opinion impact eating disorders in early adulthood? Does this vary by sex and family wealth?		As perceived peer approval increased, girls' and boys' body satisfaction increased. For girls, increasing body satisfaction was associated with a decrease in eating disorders. Boys risk of dieting varied by household wealth.

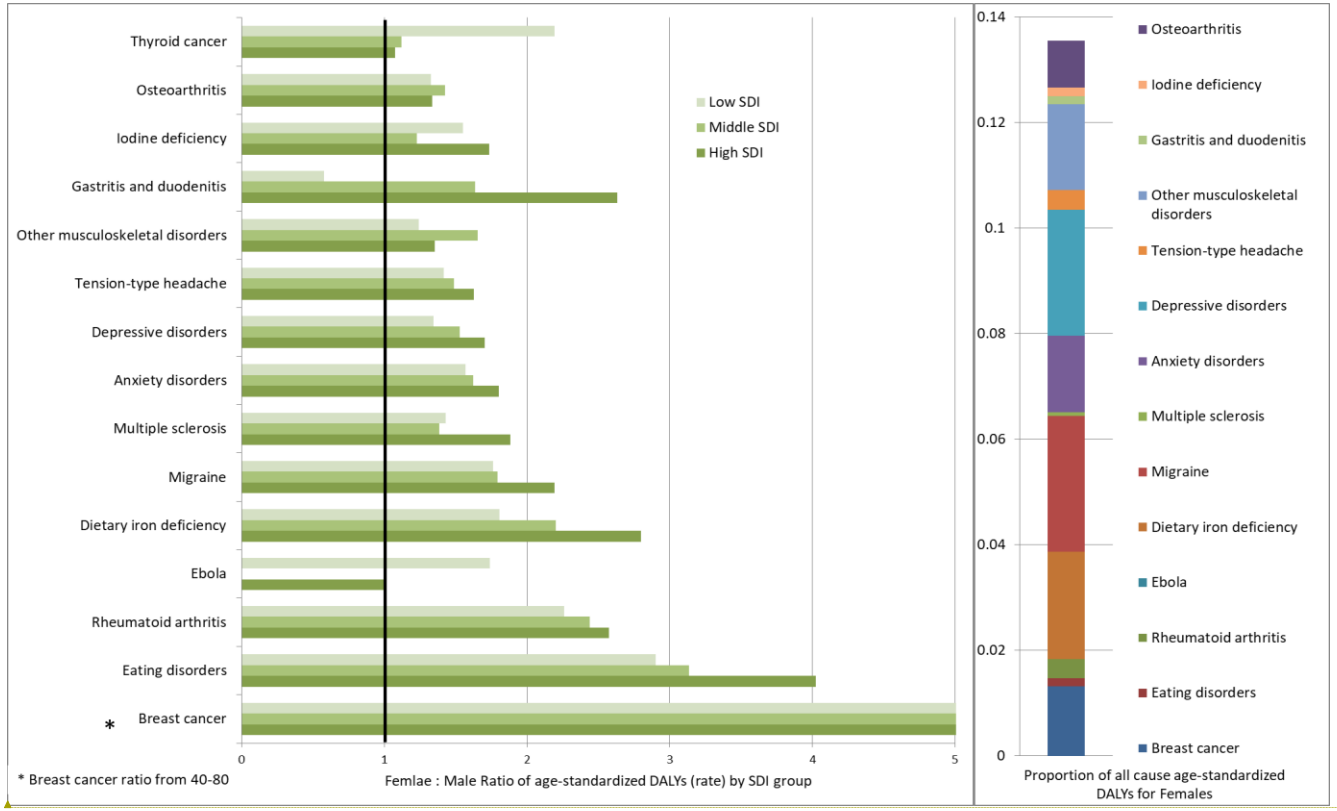
Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 3 Parental influence and mental health	Gendered health behaviours	Pelotas Birth Cohort, Brazil, 1993	Male (n=~~1245~~1113) and female (n=~~1419~~1309) youth, 11-18 y, with normal BMI at 11 y	Individual-level perception of parent's opinion of their weight	Do early adolescent perceptions of parents' opinion impact mental health in later adolescence? Does this vary by sex?		Among girls, but not boys, body dissatisfaction (feeling fatter than ideal) was associated with worse mental health outcomes when they thought their parents' opinion was also that they were fatter than ideal.
# 4 School grade peer influence and health	Gendered health behaviours	Add Health, USA, 1994-95	Male and female youth, 11-18 y (n=20,745)	Median gender normativity score of same-sex school peers (see Appendix 4 for details)	Does individual non-conformity with school peers' gender normativity impact health? Does this vary by sex and direction of non-conformity (more 'masculine' or 'feminine' than same-sex peers)?		For both girls and boys, increasing gender non-conformity with same-sex peers in either direction (i.e.: more 'masculine' or more 'feminine') was associated with increased risk for multiple health and behaviour outcomes.

Case Study	Gendered pathways[a]	Data Source	Population	Gender norm proxy	Research Questions	Diagram[b]	Results: Norm-Health association
# 5 Premarital sex and HIV status in Zambia	Gendered power disparities	DHS, Zambia, 2007	Female (n=1669) and male (n=1285) youth, 15-24 y, who ever had sex	Cluster-level (urban/rural region) average of male and female adult (25-49 y) attitudes about young people engaging in premarital sex.	Does community-level non-conformity with norms for premarital sex impact adolescent risk for HIV acquisition? Does this vary by sex?		In regions were most adults disapprove of premarital sex (and yet have premarital sex), sexually-active girls, but not boys, are at higher risk of positive HIV status
# 6 Female labour force participation (FLFP) and IPV in Nigeria	Gendered power disparities	VACS, Nigeria, 2014	13-24 y, females, (n=1633)	Indicator of community with a low % of women working outside the home (FLFP-low)	Does individual transgression of gender norms related to FLFP in low-FLFP communities impact a young women's risk of experiencing IPV?		Women who work outside the home experience higher rates of IPV than women who don't, but only in communities where working outside the home is not the norm.

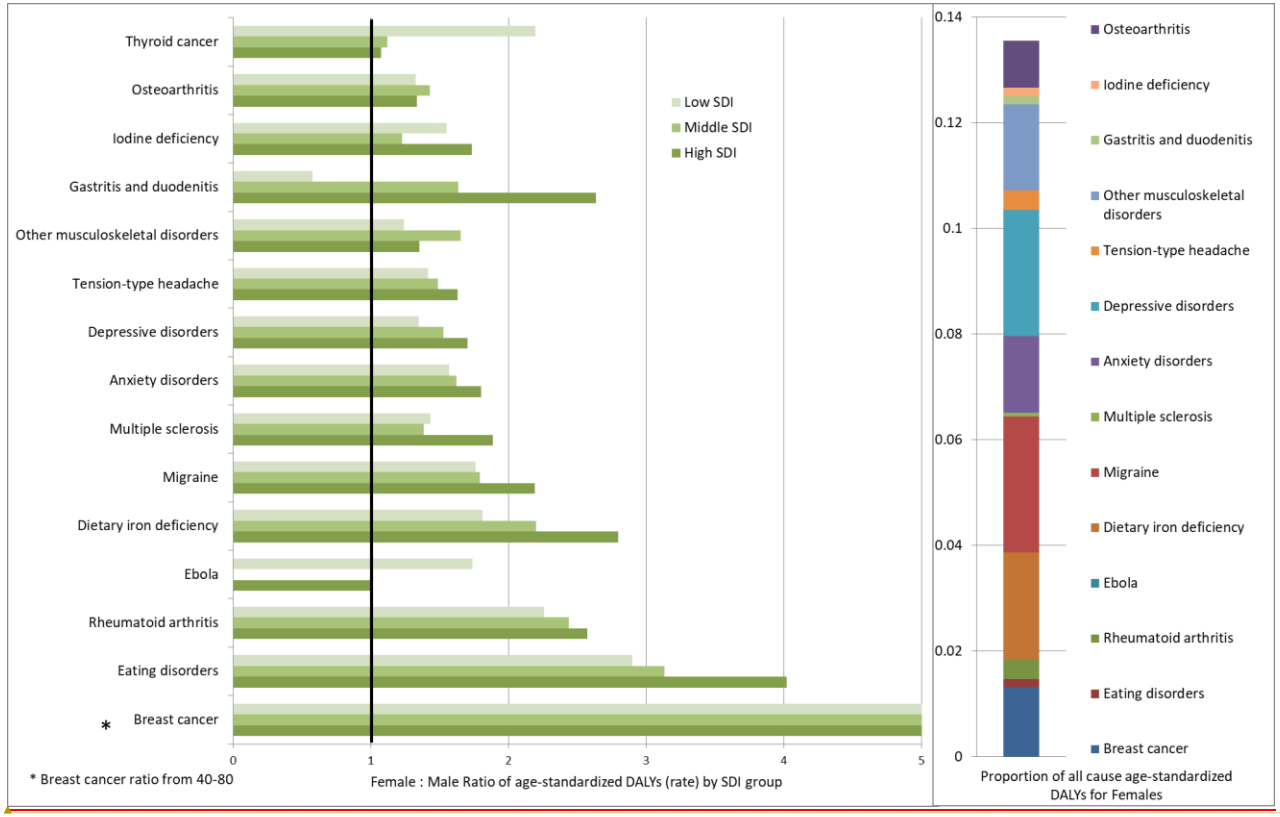
[a] The gendered pathways provide a conceptual link to the gender system and health framework presented in Heise and Greene.[10]

[b] The diagrams reflect the hypotheses we aimed to test and indicate a temporal causal direction. However, most of the data are cross-sectional and insufficient to determine causality.

Figure 1a: Female: Male ratio of age-standardized DALYs for low, middle, and high Sociodemographic Index (SDI)[a] groups (excluding low-middle and middle-high SDI countries for ease of data visualization)



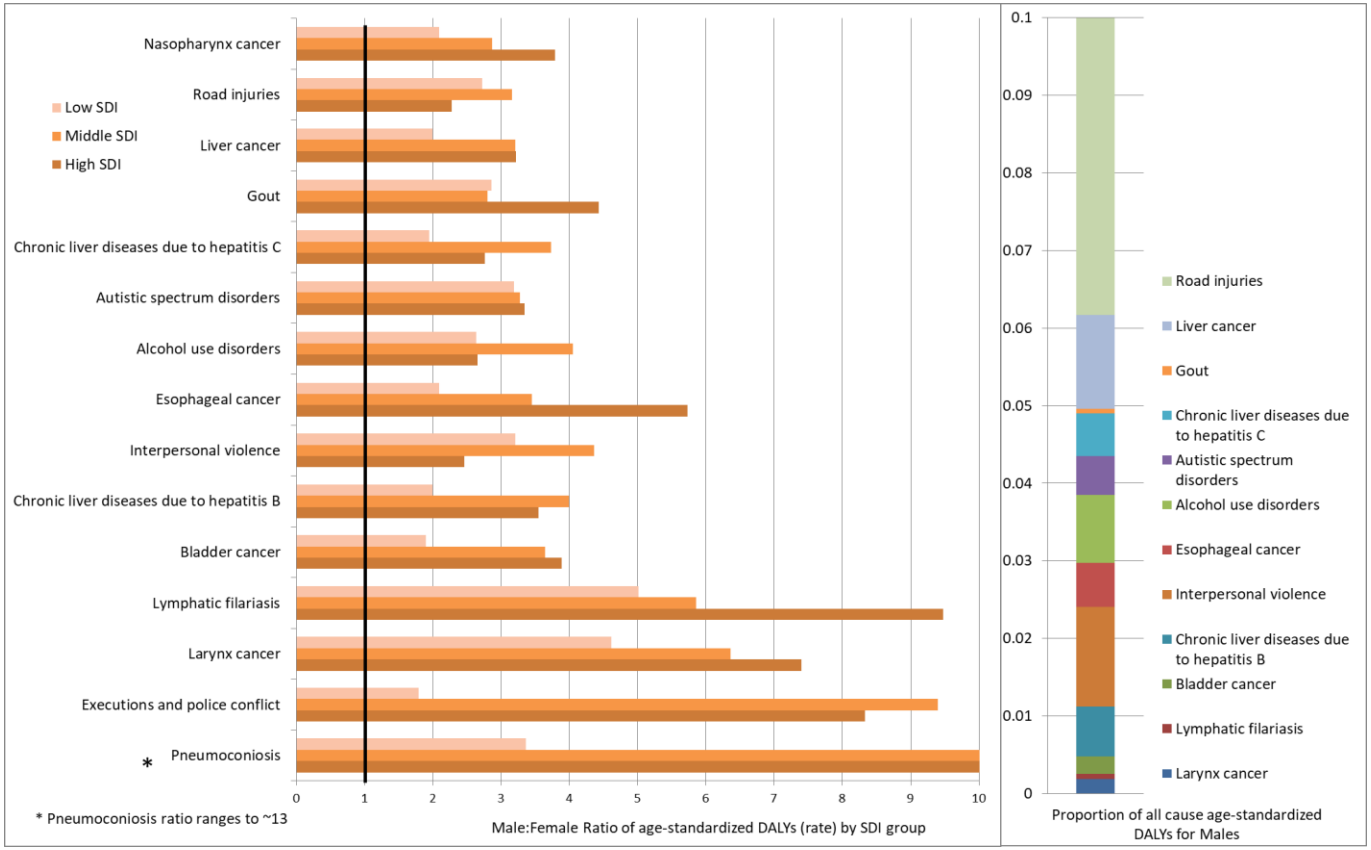
Formatted: Font: (Default) Times New Roman



[a] SDI is comprised of: average income per person, educational attainment, and total fertility rate.

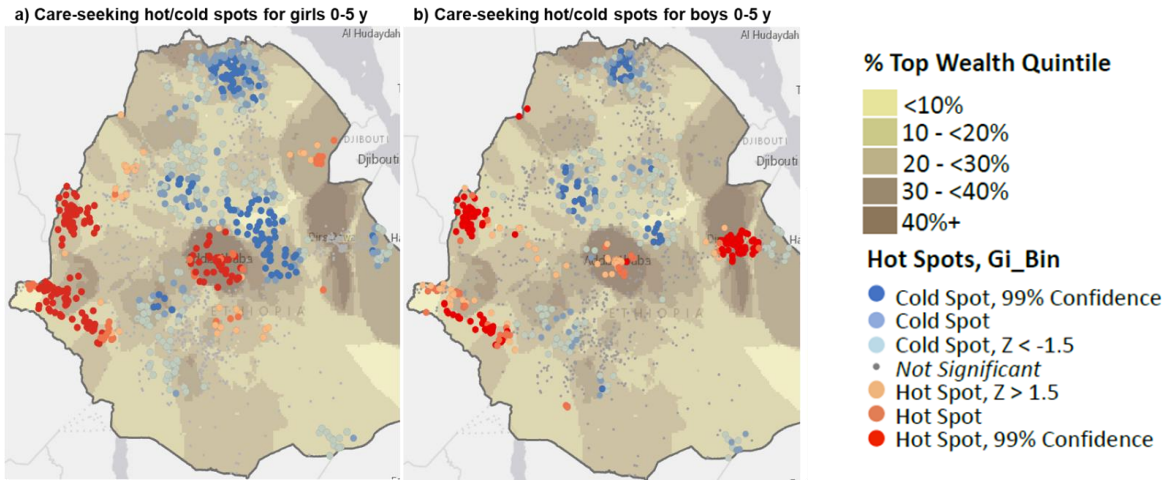
Formatted: Font: (Default) Times New Roman

Figure 1b: Male: Female ratio of age-standardized DALYs for low, middle, and high Sociodemographic Index (SDI)[a] groups (excluding low-middle and middle-high SDI countries for ease of data visualization)



[a] SDI is comprised of: average income per person, educational attainment, and total fertility rate.

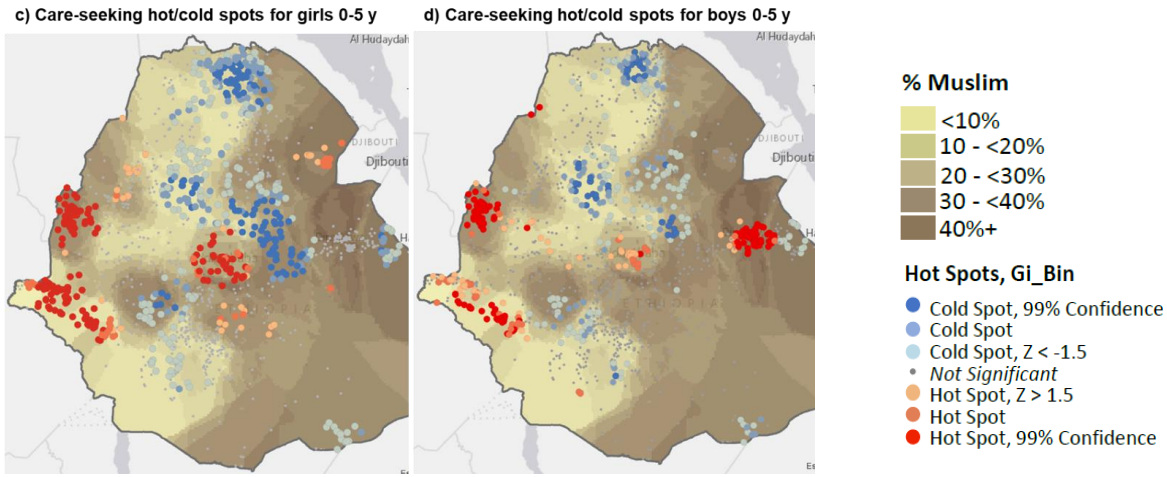
Figure 2a and 2b: Care-seeking hot/cold spots for girls (a) and boys (b) in Ethiopia by %wealthy households



Hot spots (red) and cold spots (blue) are clusters of communities with significantly higher and lower care-seeking than the national average, respectively, for girls (a) and boys (b) separately (see Appendix 1 for details). Maps are overlaid with the spatial distribution of the percentage in the communities of top wealth quintile households (for the country). The spatial distribution is displayed using kriging, a method for interpolating spatial data.[92]

~~Marginal estimates for the probability of a care-seeking hotspot are plotted at 0, 25, 50, 75, and 100% of top wealth quintile households, separately for girls (dark grey lines) and boys (light grey lines). Logistic regression models with robust standard errors clustered around nearest neighbour clusters (i.e., groups of villages in close proximity with similar z-scores) and urban status were used to predict hot spots by wealth, adjusted for % Muslim, ethnicity, parental education, and % of children vaccinated.~~

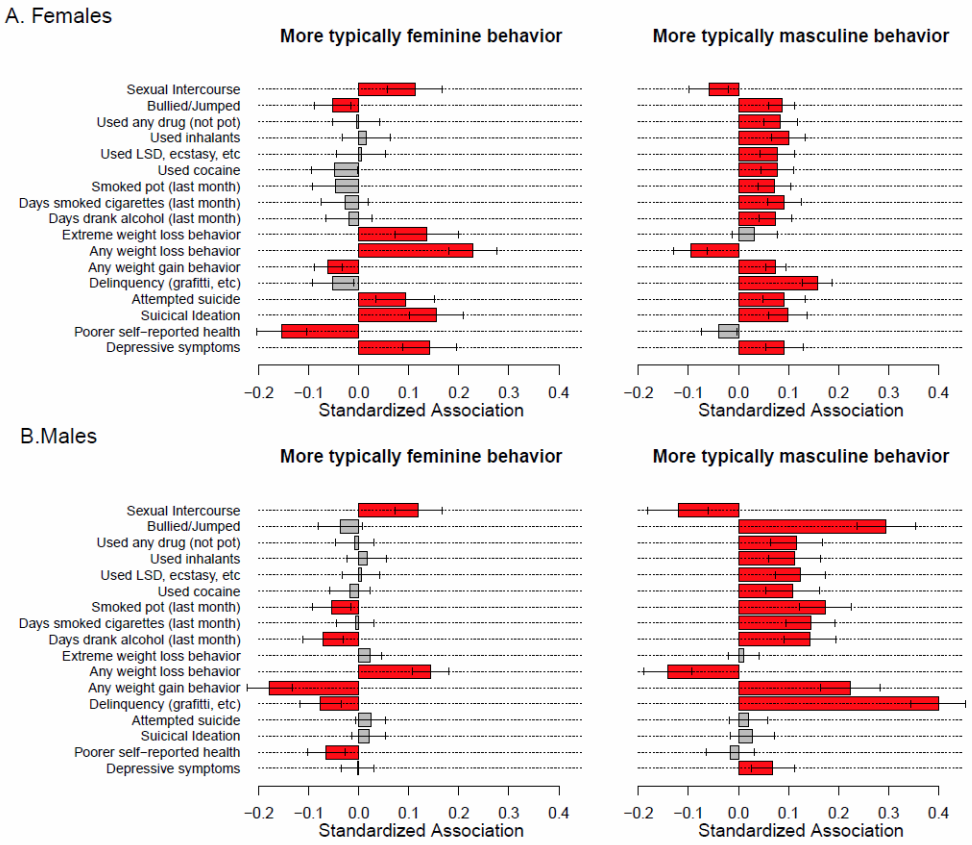
Figure 2c and 2d: Care-seeking hot/cold spots for girls (c) and boys (d) in Ethiopia by %Muslim households



Hot spots (red) and cold spots (blue) are clusters of communities with significantly higher and lower care-seeking than the national average, respectively, for girls (c) and boys (d) separately (see Appendix 1 for details). Maps are overlaid with the spatial distribution of the percentage in the communities of Muslim households. The spatial distribution is displayed using kriging, a method for interpolating spatial data.[92]

~~Marginal estimates for the probability of a care-seeking hotspot are plotted at 0, 25, 50, 75, and 100% of Muslim households, separately for girls (dark grey lines) and boys (light grey lines). Logistic regression models with robust standard errors clustered around nearest neighbour clusters (i.e., groups of villages in close proximity with similar z-scores) and urban status were used to predict hot spots by religion, adjusted for wealth, ethnicity, parental education, and % of children vaccinated.~~

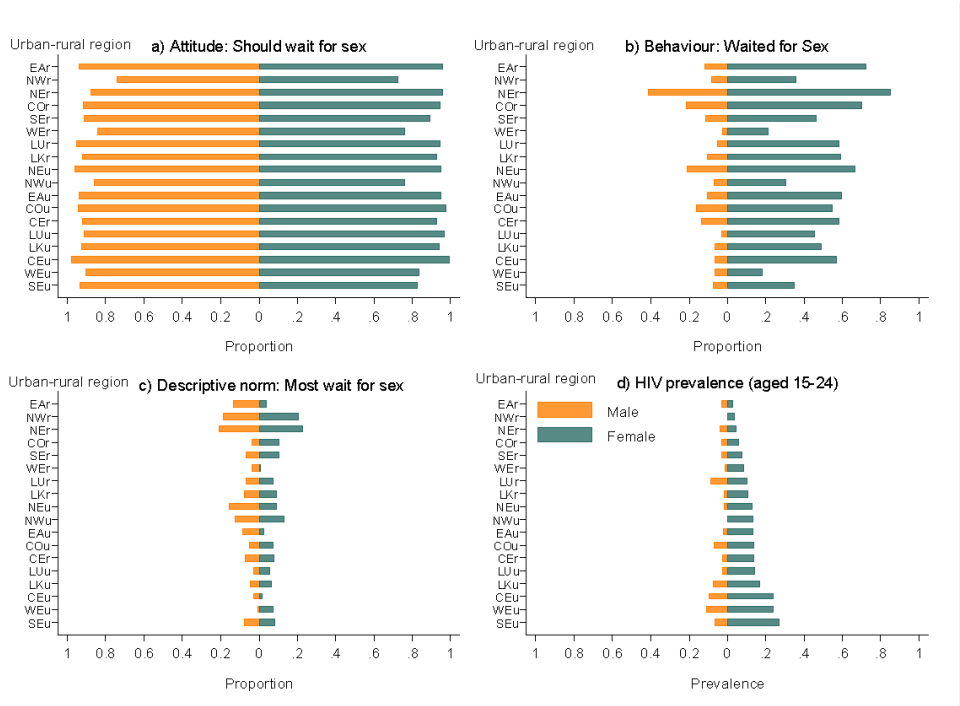
Figure 3: Estimated effects of positive and negative differences between an individual's estimated gender normativity and the median normativity of same-sex peers on health outcomes and health-related behaviours among US students, by sex.



The exposure of interest was gender norms non-conformity, or the difference between an individual's estimated gender normativity and the median of their same-sex school peers. Regressions are sex-stratified piecewise linear regressions (knot at zero) with separate effect estimates for more typically feminine and more typically masculine behaviours compared to the median of their school, controlling for an individual's own gender normativity, birth year, race, and school fixed effects. Effect estimates are

standardised so that the magnitudes can be compared across outcomes. For example, a 1 SD increase in the difference (or non-conformity) measure is associated with a 0.399 SD increase in delinquent behaviour among boys. Error bars represent 95% confidence intervals. Bars are coloured red if they are significant at the 0.01 (0.05/5) level for an appropriate Bonferroni correction based on a parallel analysis of the outcomes in the full sample, suggesting that there are 5 components.

Figure 4: Sex differentials in the proportion of adult (men and women, aged 25-49 years) for a) attitudes, b) behaviours, c) descriptive norms towards premarital sex, and d) HIV prevalence among youth (aged 15-24 years) by urban-rural regions[a] in Zambia in 2007[b]



[a]Regional codes: Central "CE~~r~~", Copperbelt "CO", Eastern "EA", Luapula "LU", Lusaka "LK", Northern "NE", Northwestern "NW", Southern "SE", Western "WE". The subscripts "u" and "r" stand for urban or rural region, respectively.

[b]Authors' estimates with information from 2007 ZDHS.

Aggregated responses were sex-stratified: men's responses about men's attitudes/behaviours and women's responses about women's attitudes/behaviours.

Web Appendix - clean

Click here to download Web Appendix: Lancet_Paper_2_appendices_17Mar19_ALL-CLEAN.docx

Web Appendix - track changes

Click here to download Web Appendix: Lancet_Paper_2_appendices_17Mar19_ALL-TRACKED.docx

Responses to Re-Reviews - Paper 2 of *The Lancet* Series on Gender Equality, Norms and Health

We wish to thank reviewers 3 and 8 for their rapid turn-around and follow-up on our previous revision. Below are our responses to their subsequent comments.

Reviewer # 3

Thank you for the chance to have another look at this interesting paper. The authors have addressed most of my comments and the paper is much improved. However, I still have a few suggestions for improvements:

Given recent discussions on p-values, I would recommend that the authors refrain from reporting their results as "marginally significant" (see examples and references below). Instead, they could reflect more carefully on the actual effect sizes and confidence intervals with a view to the quality of the analyzed data.

- *Line 227: "Differential care-seeking hot spots favouring boys had a marginally significant association with mostly wealthy (>50% of households) communities (OR=2·56, 95% CI 0·92, 228 7·12; p-value 0·071)".*
- *Line 233: "Our findings, however, are consistent with reports of son preference in other contexts, although the association with higher wealth was only marginally significant".*
- *Line 273: "For boys in lower-wealth households, increased perception of peers' approval over time was associated with a marginally significant reduction in dieting scores, with a marked reversal of this trend in higher-wealth households."*
- *Line 296: "In sex-disaggregated regression, there was a marginally significant interaction between perceived parent's opinion about weight at age 11 and body dissatisfaction at age 15 (p-value =0.052)"*

Pritschet, L., Powell, D., Horne, Z. (2016). Marginally significant effects as evidence for hypotheses: Changing attitudes over four decades. Psychological Science.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., ... & Cesarini, D. (2018). Redefine statistical significance. Nature Human Behaviour, 2(1), 6.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. The American Statistician, 70(2), 129-133.

Response: We have modified each of the referenced sentences (at lines 227, 233, 273, and 296) to remove "marginally significant." We also removed the term from the results section in Appendix 3.

Minor issues:

The textboxes (Table 1) in the "Diagram" column are now empty - I'm not sure whether this is due to a glitch in my Adobe package?

Response: The text was removed when the Word version was compiled and converted to pdf in the submission process. However, this does not happen when simply saving the document as a pdf. We have added a pdf version of the table as a separate attachment.

I still think the title is overselling the message - "How gender norms link to health" would better represent this contribution.

Response: We have modified the title to "Gender norms and health: insights from global survey data."

There is a typing error in the title for the X-axis in Figure 1a.

Response: We thank the reviewer for catching this error. It has been fixed.

The estimates presented on page 13 do not correspond with the results reported in Table A2.2 and Figure A.2.1 - I imagine that the authors have forgotten to update the estimates in the main text.

Response: We thank the reviewer for catching these inconsistencies. These have been fixed.

In line 232 in Appendix 2, "(table A2.2)" should be changed to "(table A2.3)".

Response: Thank you. We have fixed the error.

Reviewer #8

This revised version of the manuscript on gender norms has improved following the many comments from multiple reviewers. The authors have done well to address them in a short amount of time. The authors have attempted to present a balanced approach by mentioning where results are "significant", "marginally significant" and "not significant", but this does not address the point previously raised. There is a growing movement away from over-reliance on p-values and use of the term "significant" altogether. For example see https://rdcu.be/bpZwD Effect sizes, confidence intervals and consistency of the direction of effect are more informative aspects to highlight.

Response: We thank the reviewer for acknowledging our effort to address the comments with a short turn around. We appreciate this additional opportunity to revisit some of the reviewer recommendations and the changes we made. Regarding the reliance on p-values, Reviewer 3 expressed a similar concern and requested that we remove the term "marginally significant" entirely, which we have done. We also reviewed our use of "significant" and "not significant" throughout. In a few sentences, we removed the reference to significance (if it was not important) or balanced it with a comment on the magnitude of the association.

The presentation of case studies 1 & 2 might be improved further with elaboration below. For the other case studies I offer only minor or no further suggestions.

****Case study 1*
The analysis has changed from the original version of the manuscript, with covariates included in a different form to the previous version. There is some consistency with the previous version which also used similar cut-points to define the covariates. However there are still some issues with this case study explained below.

It is not clear to me if it was decided not to test any interactions, or whether the interaction was tested and not found to show something. What was the original statistical plan regarding testing for effect modification?

Response: The statistical plan for all case studies was to test, when feasible, how gender norms interact with other social factors to influence health. In this way, we aimed to test the theory of "intersectionality" in which one's social position is influenced by inter-related inequalities based on

social class, ethnicity, and gender, etc. (as described in Paper 1 in the series). In the Ethiopia case study, we tested how wealth and religion were related to the gender norms proxy of differential care-seeking. We saw a unique opportunity with this case study to test the interaction effect of these two factors on gender norms.

In the original version of the paper, we presented the interaction term for high/low % wealthy and high/low % Muslim communities. However, in response to the two statistical reviews we received about this finding, we revisited our original analyses and asked whether the finding was robust to different formulations (e.g., with the inclusion of the main effect terms, or the use of continuous vs. binary variables for the predictors). In continuous formulation of both factors, including main effects terms, the interaction term was significantly associated with differential CS hot spots. Because most communities are mostly rich or not, and mostly Muslim or not (i.e., bimodal), it is not clear that the interpretation of the parameter estimates based on the continuous variables is meaningful. However, an estimate of the interaction effect on differential CS hot spots could not be obtained in a model that included the main effects and interaction terms using all binary indicators. Low % wealthy and low % Muslim perfectly predicted not being a hot spot (no communities were hot spots for differential CS and low % wealthy and low % Muslim). We concluded that the previously reported interaction effect (without main terms) on differential care-seeking hot spots was likely driven by the large effect of religion (93 of the 107 hot spots are in Muslim majority communities). Therefore, we removed the finding from the main paper.

In this revision, we ran tests of interaction on CS hot spots for boys and for girls separately with main and interaction terms and followed the process in the recommended Knol (2012) paper to evaluate interaction on both an additive and multiplicative scale (see new Tables A1.5 and A1.6). We thought that if there were interactions in the sex-specific models, this would provide some evidence of an interaction for the differential CS. The tests of interaction were positive for boys and negative for girls on the additive scale, though neither test was significant, and both demonstrated negative interaction on the multiplicative scale and were also not significant. We have added a comment in the text that we tested for an interaction of percentage wealthy and Muslim households on care-seeking, but that evidence for an interaction was not found.

I am generally suspicious of presented results where OR are so high as to whether the model is a good fit, were goodness of fit statistics looked at? Although I do note that the crude OR is also high (15.5 by my calculations).

Response: The reviewer is correct that the crude OR is 15.5 (we have added a new table A1.3 that allows the reader to calculate this easily). We were also surprised by the magnitude of the OR for Muslim majority communities. The association was insensitive to cut-point of the z-statistic (1 or 2, instead of 1.5) and the model explains about 31% of the variance in differential CS hot spots. However, ORs can over-estimate the RR, so we estimated the RR using the conditional means of being a hot spot given covariates from the adjusted logistic regression. The RR was on the order of 8 to 9, still a a very large effect estimate. We have added the limitation of ORs over-estimating RRs in the results section of the appendix. Also, we reverted back to the continuous forms originally used for Somali and Christian, as we do not discuss the coefficients on these factors and the continuous form explains more of the variation in the outcome. This change resulted in a reduction in the adjusted OR to 18.

********Case study 1 appendix*
"The second step was to use linear regression models with the dichotomized spatially-weighted z-scores as the" should this sentence say "logistic regression"?

Response: The reviewer is correct. We have changed this sentence to say “logistic regression.”

I still cannot get my head around what it means to have differential care seeking based on what is written here in the appendix. It might help to include a numeric example for one cluster or to use language more similar to your response to reviewers where it was much clearer.

Response: We used some of the language in the previous response to reviewers and expanded the description of the outcome in the appendix. We hope that the reviewer finds this description clear.

I find the rows in Table A1.2 for care-seeking % very confusing, how should these proportions be interpreted, and why are the sample sizes different to other variables? If a hot spot is defined as a z-statistic>1.5 I don't understand why the mean z-statistic for girls in hot spots is 0.08, and indeed why its mean is less than the non-hot spots?

Response: We thank the reviewer for raising these questions with respect to Table A1.2. Regarding the second question, the rows for boys and girls CS z-statistics were reversed and this has now been fixed. Regarding the first question, the care-seeking percentages are the average percentage of girls, boys, or combined boys and girls (total) under 5 in a community who were ill in the previous two weeks and who received medical care from a certified medical practitioner. The % care-seeking for boys minus girls is the average difference in percentages of boys and girls who were ill and received care in a community. The sample size, N, for these factors is lower than that for the hot spot sub-groups, because some communities did not have any ill girls or boys or both in the 2 weeks prior to the survey. For example, in 81 differential CS hot spot communities (out of 107), an average of 31% of girls under 5 who were ill received medical care. It is important to know that all communities received a z-statistic for being a hot spot or not, even if they had no ill children. This is because of the smoothing routine performed by the geospatial analysis that interpolates across geospatially proximal communities. We have added the above clarification to the appendix and the legends of the tables.

Table A1.1 Some numbers are in bold with no explanation, assume due to low p-value, but these p-values will not adjust for clustering, so are they valid?

Response: The bolded numbers were for correlations above 0.7, suggesting that those variables may be subject to variance inflation due to collinearity. We have removed the bold.

Table A1.2 statistical tests are mentioned in the text but not signposted in the text to the table, this might be added within the modelling section, although these tests are not adjusted for clustering but the later regression models are; are these p-values valid?

Response: The statistics in Table A1.1 are community level comparisons, but these communities are clustered into hot and cold spots with the geospatial analysis. We have repeated the statistical comparisons, adjusting for the geospatial clustering and updated the table and legend to reflect these results. We have also retained the continuous forms of all percentage variables in this table, only dichotomizing % wealthy and % Muslim in the regressions, given their largely bimodal distributions, for a more meaningful interpretation of the coefficients (as discussed above).

In the footnote to table A1.3 "Omitted automatically by model due to collinearity" please add variable collinear with.

Response: This was another situation of perfect prediction (no hot spots for girls in high % Somali

villages). We resolved the problem when we reverted to the continuous form of the variable (as discussed above).

In the adjusted model, it would be better to centre the remaining continuous variables in the model. A better presentation than table A1.3 and A1.2 would include the N with/N without outcome in the same table as the OR, CI and p-values and include the reference level in the table. This would be a similar presentation to the previously mentioned article table 1 for A=0
https://www.ncbi.nlm.nih.gov/pubmed/22253321

Response: We have centred the continuous variables in the adjusted models, as the reviewer suggested. We also have created new tables based on the recommended article for presenting tests of interaction (see previous notes).

********Case study 2*
I do agree that change in one variable being modelled against the change in the other is a correct thing to do. However, it would be useful to additionally adjust for the baseline covariate as the values change may take do still depend on where an individual started, so an ANCOVA analysis is still more powerful for a change outcome, than an unadjusted analysis.

Response: We agree that adjustment for the value of outcome at baseline is useful in the case when it is a suspected confounder, i.e., that it is independently associated with both the follow-up measurement of the outcome and the predictor. While it is possible that the baseline value of the outcome is linked with the change in the outcome variable between baseline and follow-up, we do not have reason to suspect that the baseline measure of the outcome variable is independently associated with the change in predictor variable between baseline and follow-up for reasons other than that it is correlated with the baseline value of the predictor, which is the association we are seeking to investigate in a longitudinal way. In this case, adjusting for the baseline measure of the outcome could lead to over-adjustment and attenuation of the associations. Also, inadvertently adjusting for the baseline value does not necessarily translate into unbiased results when there is measurement error and unmeasured confounders, as is likely in our case (see for example the article by Lepage B et al, Epidemiology. 26(1):122–129, JAN 2015: (https://insights.ovid.com/pubmed?pmid=25401453). When we adjust for baseline values we cannot know if any different results are because we removed a potential confounding effect, introduced a source of bias, or over-adjusted. Nevertheless, we included a sensitivity analysis where we adjust for the baseline values, and the conclusions about indirect effects remain essentially the same (i.e. conclusions related to all outcomes in boys and all outcomes in girls remain the same except for oral control scores where the indirect effect is attenuated to the extent that it loses statistical significance). The effects for the dieting and bulimia scores outcomes among girls are slightly attenuated but are still statistically significant and in the same direction. We have included this finding in the results for Appendix 2 and added a corresponding figure A2.3.

The presented tables are somewhat different to the first version of the manuscript. Table A2.1 is confused by the additional of "n=331" and "n=277". The table is signposted as being the missing data for key variables, but those sample sizes are for the complete case data, should that text be omitted from this table?

Response: Yes, we have removed that text.

There is a big size difference between the full sample and the complete case analysis sample, what were the factors associated with missingness (worthwhile mentioning in the results section of the appendix for the reader to make judgements regarding potential selection bias), and are these

adjusted for? It is reassuring at least that the same message comes out from the previous version of the manuscript.

Response: The size difference between the complete case analysis and the full sample is due to missingness of several variables, including the outcome variables at age 22 follow-up. We have checked the factors associated with missingness of age 22 outcome variable and included a statement to the effect in the Appendix (results section). Those with complete data for the outcome variables were more likely to come from wealthier households. Household wealth was included in the models. Importantly, we tested the results with multiple imputation, reassuringly finding very similar results.

****case study 3*
The revised presentation is clearer. Might consider adding a sentence to the results as for case study 2 with summary of factors associated with missingness.

Response: Thank you. Among the cohort subset used in the analysis (those who had normal BMI at age 11), there was no missingness in the outcome. Even covariates had very small percentage of missing values, with a maximum of 5% among girls and 6% among boys in BMI change between 11 and 15 years, as evident in table A3.1. We now note that in the results section of Appendix 3.

***case study 4*
The revised presentation is clear.

Regarding table A4.5 where covariates are shown as (Y/N) I query if it useful to show the standard deviation of the proportion. Typically the number and % (to 1 decimal place) are presented for binary variables in a descriptive table (e.g. as in table A5.1).

Response: We have updated table A4.5 to match the presentation of Table A5.1.

Given the other case studies have used complete case analysis and presented tables with missing values, might it be appropriate to take a consistent approach across the manuscript and do the same for case study 4?

Response: For case study 4, our concern with performing a complete case analysis is the possibility of introducing bias by restricting the regressions to cases with complete outcome data that are not all related. However, with the exception of the two questions on sex, the sample sizes do not vary substantially. Therefore, as a sensitivity check, we performed a complete case analysis for all outcomes excluding the two questions related to sex. The magnitudes of all the effects remained comparable. Only a decreased risk of "smoking pot" for more feminine males changed significance (became non-significant). Therefore, we prefer to retain the maximum sample for each outcome.

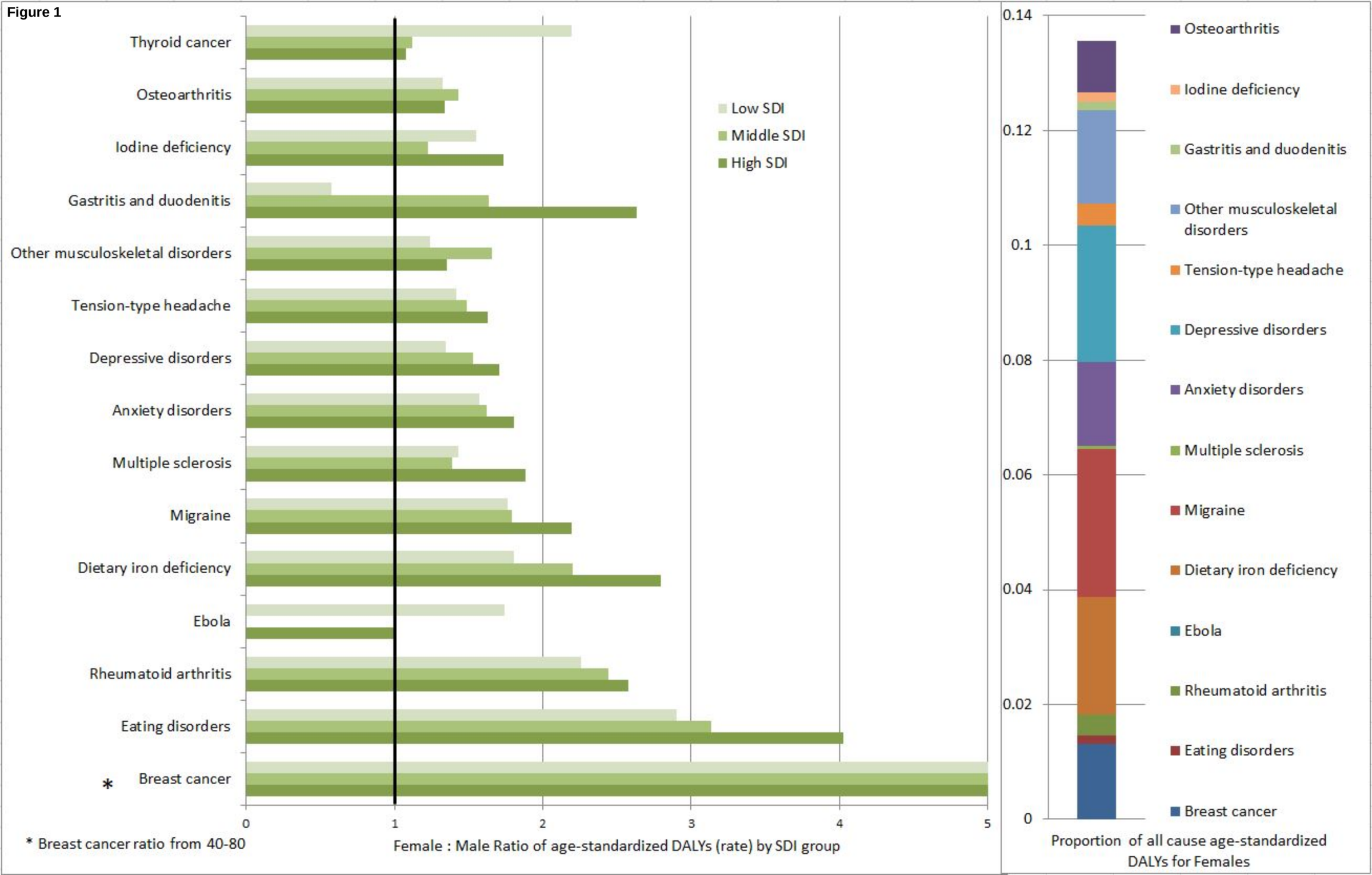
****case study 5*
Should the continuous variables have RR which represent a 1 unit change, or would a higher unit change of 5, 10 or 20 be more appropriate given the closeness of the RR to 1? The units are not always clear in the table (A5.2 &A5.3), for example education is it in years or is it some other unit?

Response: We thank the reviewer for this suggestion. We have updated tables A5.2 and A5.3 to represent the RR for a 10% unit change. The text in the main paper and appendix have been updated to correspond to these values.

***case study 6*
The revisions are clear.

Response: Thank you!

Figure 1


Thyroid cancer
Osteoarthritis
Iodine deficiency
Gastritis and duodenitis
Other musculoskeletal disorders
Tension-type headache
Depressive disorders
Anxiety disorders
Multiple sclerosis
Migraine
Dietary iron deficiency
Ebola
Rheumatoid arthritis
Eating disorders
* Breast cancer
Low SDI
Middle SDI
High SDI
0
1
2
3
4
5
* Breast cancer ratio from 40-80
Female : Male Ratio of age-standardized DALYs (rate) by SDI group
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
Osteoarthritis
Iodine deficiency
Gastritis and duodenitis
Other musculoskeletal disorders
Tension-type headache
Depressive disorders
Anxiety disorders
Multiple sclerosis
Migraine
Dietary iron deficiency
Ebola
Rheumatoid arthritis
Eating disorders
Breast cancer
Proportion of all cause age-standardized DALYs for Females

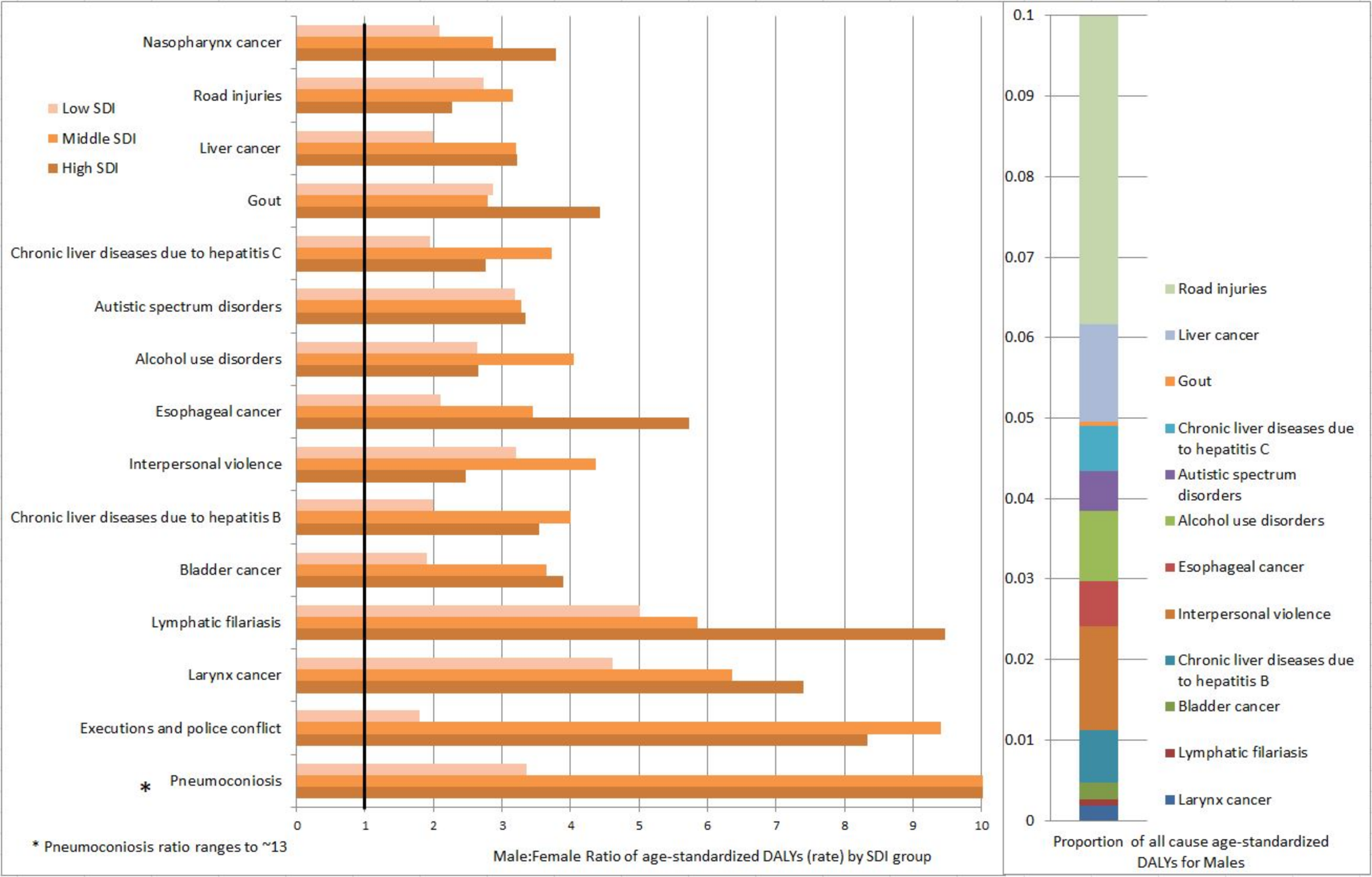
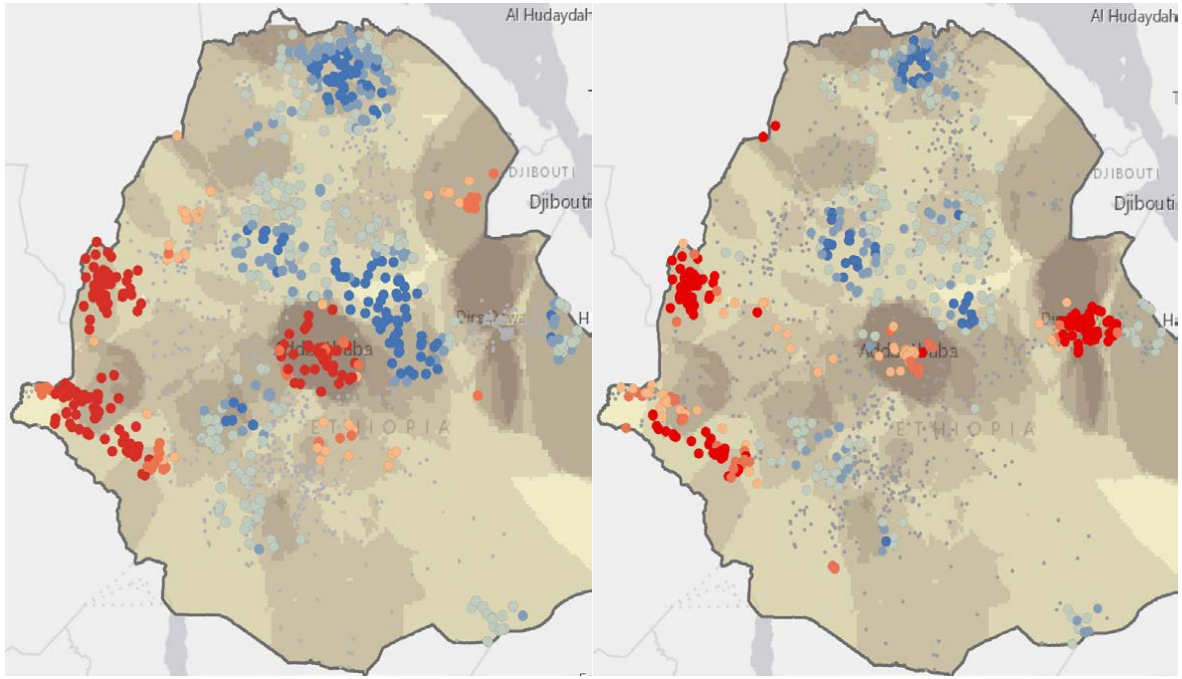

Low SDI
Middle SDI
High SDI
Nasopharynx cancer
Road injuries
Liver cancer
Gout
Chronic liver diseases due to hepatitis C
Autistic spectrum disorders
Alcohol use disorders
Esophageal cancer
Interpersonal violence
Chronic liver diseases due to hepatitis B
Bladder cancer
Lymphatic filariasis
Larynx cancer
Executions and police conflict
* Pneumoconiosis
0
1
2
3
4
5
6
7
8
9
10
* Pneumoconiosis ratio ranges to ~13
Male:Female Ratio of age-standardized DALYs (rate) by SDI group
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
Road injuries
Liver cancer
Gout
Chronic liver diseases due to hepatitis C
Autistic spectrum disorders
Alcohol use disorders
Esophageal cancer
Interpersonal violence
Chronic liver diseases due to hepatitis B
Bladder cancer
Lymphatic filariasis
Larynx cancer
Proportion of all cause age-standardized DALYs for Males

Figure 2

a) Care-Seeking hot/cold spots for girls 0-5 y

b) Care-seeking hot/cold spots for boys 0-5 y



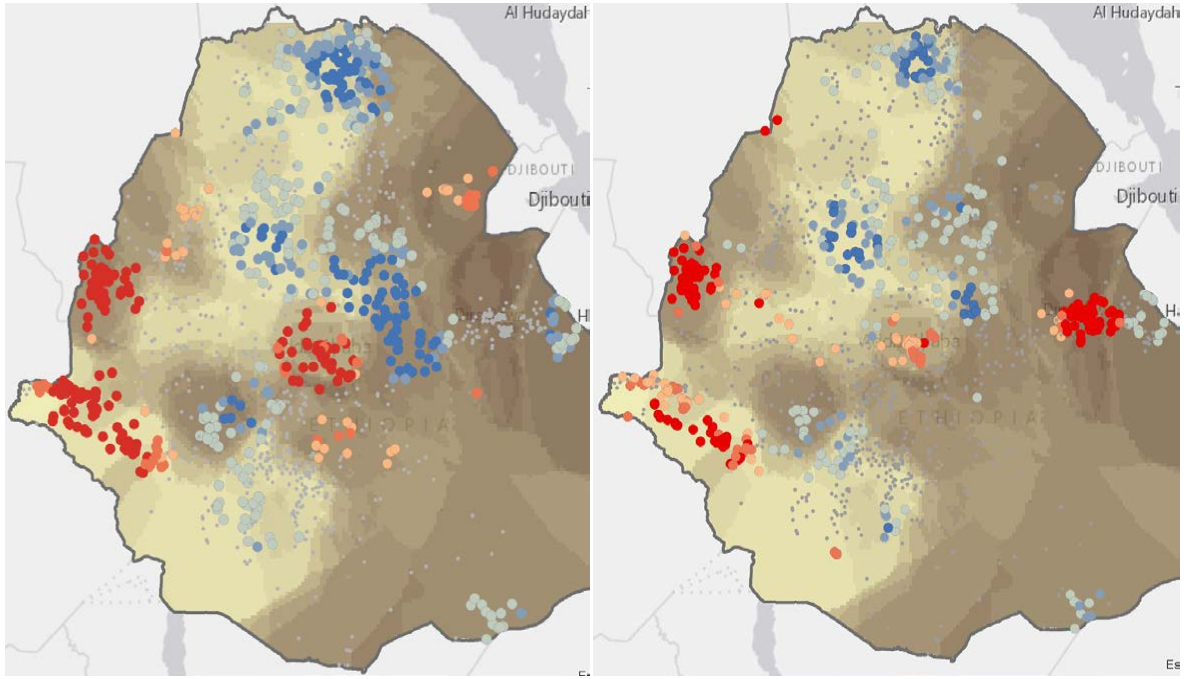
% Top Wealth Quintile

- <10%
- 10 - <20%
- 20 - <30%
- 30 - <40%
- 40%+

Hot Spots, Gi_Bin

- Cold Spot, 99% Confidence
- Cold Spot
- Cold Spot, Z < -1.5
- *Not Significant*
- Hot Spot, Z > 1.5
- Hot Spot
- Hot Spot, 99% Confidence

a) Care-Seeking hot/cold spots for girls 0-5 y b) Care-seeking hot/cold spots for boys 0-5 y



% Muslim

- <10%
- 10 - <20%
- 20 - <30%
- 30 - <40%
- 40%+

Hot Spots, Gi_Bin

- Cold Spot, 99% Confidence
- Cold Spot
- Cold Spot, Z < -1.5
- *Not Significant*
- Hot Spot, Z > 1.5
- Hot Spot
- Hot Spot, 99% Confidence

Figure 3

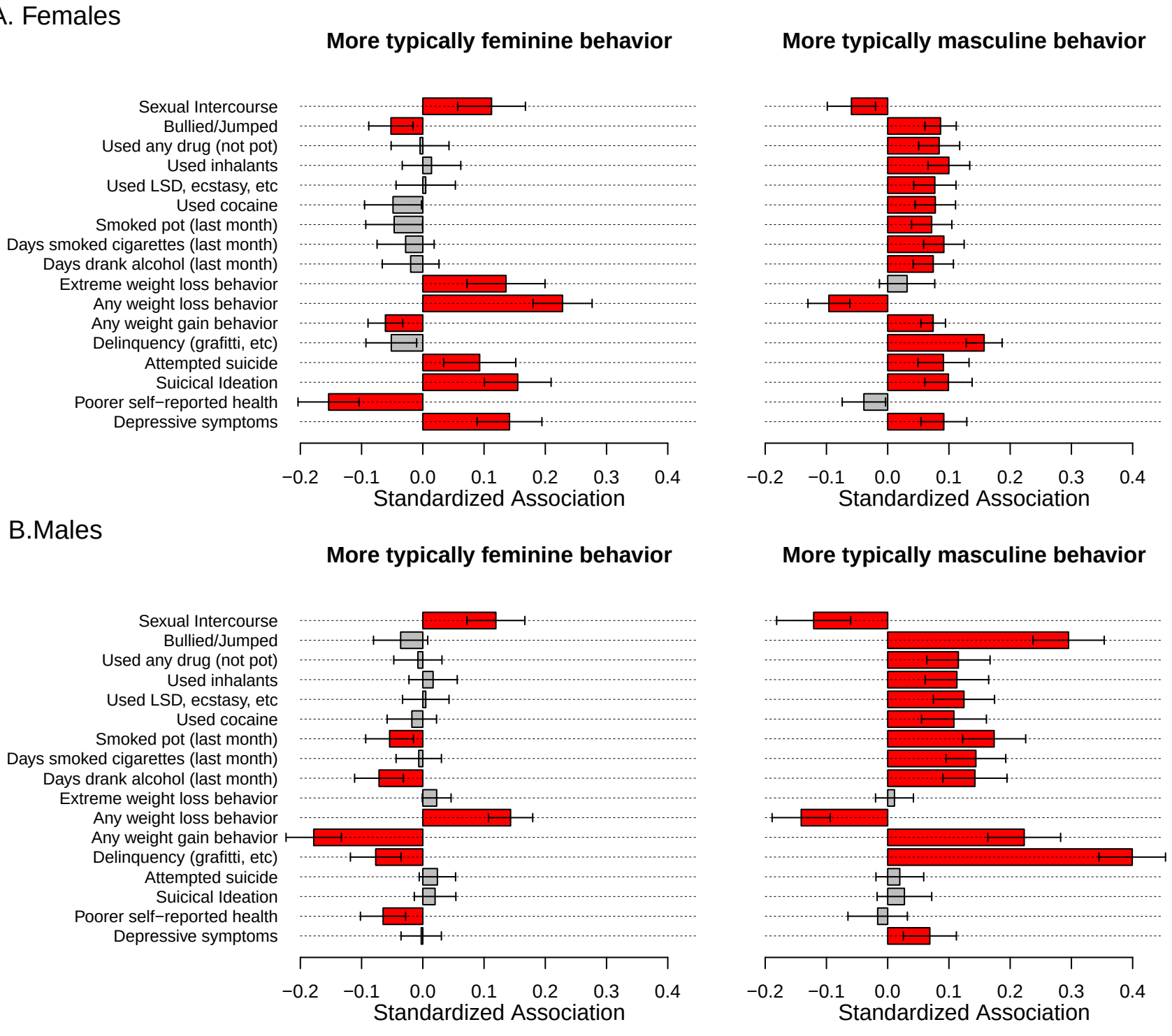
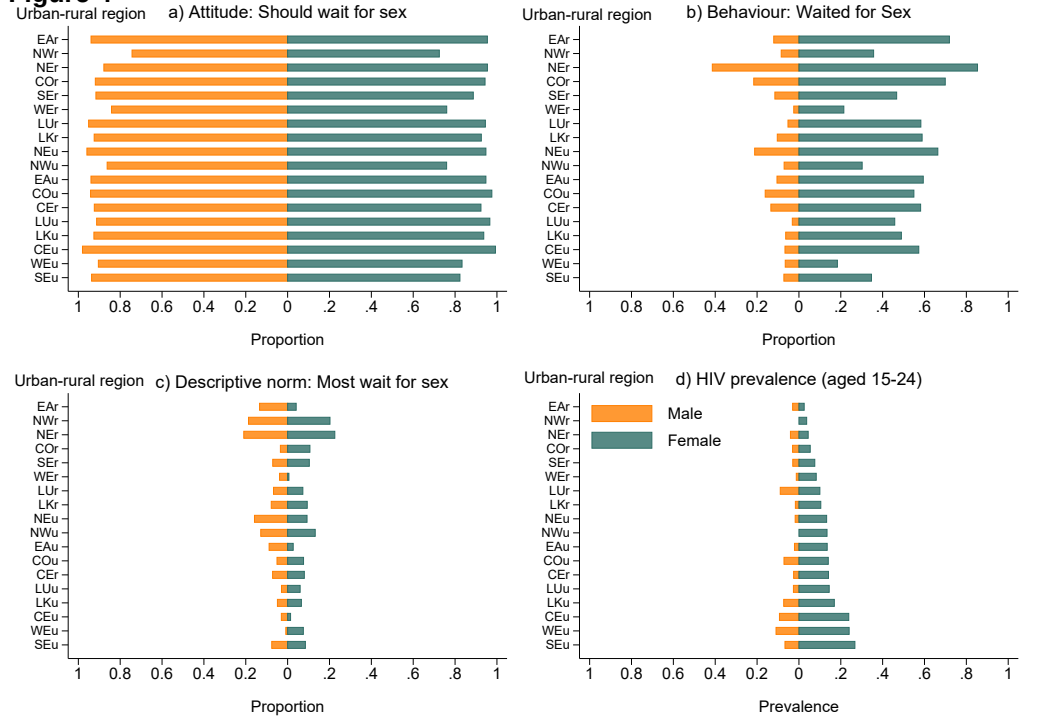

A. Females
More typically feminine behavior
More typically masculine behavior
Sexual Intercourse
Bullied/Jumped
Used any drug (not pot)
Used inhalants
Used LSD, ecstasy, etc
Used cocaine
Smoked pot (last month)
Days smoked cigarettes (last month)
Days drank alcohol (last month)
Extreme weight loss behavior
Any weight loss behavior
Any weight gain behavior
Delinquency (grafitti, etc)
Attempted suicide
Suicical Ideation
Poorer self−reported health
Depressive symptoms
−0.2 −0.1 0.0 0.1 0.2 0.3 0.4
Standardized Association
B.Males
More typically feminine behavior
More typically masculine behavior
Sexual Intercourse
Bullied/Jumped
Used any drug (not pot)
Used inhalants
Used LSD, ecstasy, etc
Used cocaine
Smoked pot (last month)
Days smoked cigarettes (last month)
Days drank alcohol (last month)
Extreme weight loss behavior
Any weight loss behavior
Any weight gain behavior
Delinquency (grafitti, etc)
Attempted suicide
Suicical Ideation
Poorer self−reported health
Depressive symptoms
−0.2 −0.1 0.0 0.1 0.2 0.3 0.4
Standardized Association

Figure 4


Urban-rural region
a) Attitude: Should wait for sex
EAr
NWr
NEr
COr
SEr
WEr
LUr
LKr
NEu
NWu
EAu
COu
CEr
LUu
LKu
CEu
WEu
SEu
1 0.8 0.6 0.4 0.2 0 .2 .4 .6 .8 1
Proportion
Urban-rural region
b) Behaviour: Waited for Sex
EAr
NWr
NEr
COr
SEr
WEr
LUr
LKr
NEu
NWu
EAu
COu
CEr
LUu
LKu
CEu
WEu
SEu
1 0.8 0.6 0.4 0.2 0 .2 .4 .6 .8 1
Proportion
Urban-rural region
c) Descriptive norm: Most wait for sex
EAr
NWr
NEr
COr
SEr
WEr
LUr
LKr
NEu
NWu
EAu
COu
CEr
LUu
LKu
CEu
WEu
SEu
1 0.8 0.6 0.4 0.2 0 .2 .4 .6 .8 1
Proportion
Urban-rural region
d) HIV prevalence (aged 15-24)
Male
Female
EAr
NWr
NEr
COr
SEr
WEr
LUr
LKr
NEu
NWu
EAu
COu
CEr
LUu
LKu
CEu
WEu
SEu
1 0.8 0.6 0.4 0.2 0 .2 .4 .6 .8 1
Prevalence