WORLD ECONOMIC FORUM

In collaboration with

McKinsey Health Institute

Technical appendix

Closing the women's health gap: A $1 trillion opportunity to improve lives and economies

January 2024

Technical appendix

This report presents the results of our investigation of women's health burden, the size of the health gap relative to men, and the potential to close that gap by addressing sources of structural or systematic bias. Building on and extending a report published by the McKinsey Global Institute in 2020,[1] the research quantified the potential impact of reducing the gap on women's health and the economic impact of that health improvement over a 20-year period to 2040. This technical appendix outlines the approach and key assumptions of that investigation.

Estimating disease reduction potential

The analysis is based on a detailed assessment of the 64 conditions that account for almost 86 percent of the female global disease burden in total (excluding injuries).

The first stage in the analysis was to assess the extent to which the current and projected disease burden (to 2040) could be addressed through existing interventions and breakthrough technologies expected to be used in clinical practice in the next 20 years. For each intervention, we identified and quantified where these therapies are known to have lower adoption rates or lower effectiveness in females than in males.

We started with the projected baseline disease burden by sex, age group, year, and country, for 195 countries. The approach consisted of three stages:

1. Identify the diseases and conditions that contribute 86 percent of the global female disease burden (excluding injuries) for in-depth analysis.
2. For the 64 conditions selected, conduct a disease-level analysis to identify the potential to improve health (and reduce the burden) through existing and breakthrough interventions and identify where there is a known sex difference in either treatment adoption (uptake) or effectiveness. Then estimate the share of the average sex-based health gap across the conditions and interventions where sex differences have been studied, and apply this difference to the conditions where sex differences have not been studied (to estimate the "unknown" health gap).
3. Extrapolate this analysis to the health burden not included in the in-depth analysis (14 percent of the global female disease burden, excluding injuries).

The analysis indicates that 44 percent of the female disease burden could be reduced with more consistent and timely delivery of proven existing interventions and breakthrough innovations. A further 6 percent could be reduced by addressing areas of inequality affecting women and girls. The remaining 50 percent of the disease burden is likely to remain unaddressed without investment in research and innovation beyond the visible pipeline of today. The focus of this report is the 6 percent that makes up the women's health gap.

Selecting diseases for in-depth analysis

The starting point for the disease impact model was the Institute for Health Metrics and Evaluation (IHME) Global Burden of Disease (GBD) data set for 2019, which categorizes all diseases in a four-level hierarchical classification system and projects the disease burden using a measurement of disability-adjusted life years, known as DALYs.[2] GBD takes into account all time lost to early death, poor health, or disability. The DALYs attributable to a disease are the sum of two parts: years of life lost (YLLs), which

[1] For the full McKinsey Global Institute report, see "Prioritizing health: A prescription for prosperity," McKinsey Global Institute, July 8, 2020.
[2] For more details, see "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," *Lancet*, October 2020, Volume 396.

measures years lost to premature mortality, and years lived with disability (YLDs), which measures the time lost to poor health or incapacity. All disease states have a weighting between 0 and 1, where 0 indicates perfect health and 1 is the worst possible health (equivalent to death). YLDs is the product of years spent with the disease and the weighting for that disease. One DALY averted can be considered equivalent to an additional one year of healthy life.

We analyzed this data set to identify the conditions accounting for 86 percent of the female global disease burden (excluding injuries), measured in DALYs, at level three (which includes 171 disease groups) and level four (to further account for gynecological diseases and maternal disorders), creating a list of 64 diseases for in-depth analysis (Exhibit 1). Conditions defined as "other" were excluded from the list of diseases for in-depth analysis.

Estimating the health improvement potential

For each condition, we estimated the potential to reduce morbidity and mortality by 2040 through greater, more consistent, and more timely use of proven interventions, considering sex differences in adoption and effectiveness of interventions to identify and quantify the women's health gap with the following process.

Step 1: Assessing modifiable risk factors associated with each condition. For each disease, we used the IHME Global Burden of Disease 2019 data set to identify the health risks associated with the condition and the proportion of the disease burden associated with each risk. For diseases with major interdependencies between risk factors—for example, cardiovascular disease and diabetes—we selected interventions that addressed multiple interdependent risk factors simultaneously.[3]

Step 2: Identifying high-impact scalable interventions. We defined health interventions as actions aimed at assessing, promoting, or improving the health of an individual or population—for example, public sanitation programs or surgical procedures—recommended by leading institutions such as the World Health Organization or national medical associations. We identified relevant clinical guidelines, systematic reviews, and medical literature for each disease to identify and categorize the interventions with the greatest potential for scalable impact, looking at both interventions with the potential to prevent the disease and interventions to treat established disease. Whenever possible, we relied on existing internationally focused evidence reviews from organizations and initiatives, including *Disease Control Priorities* (DCP3), the World Health Organization, and similar agencies. We then searched for peer-reviewed published studies investigating and reporting sex differences in either effectiveness of or access to (adoption or uptake of) the interventions identified in the first phase. The research team included medical doctors and clinical experts, who reviewed all findings. In all cases, the aim of the research was to identify a basket of cost-effective, critical interventions with wide applicability, rather than to catalog an exhaustive list of all possible treatments.

In total, we identified around 183 relevant interventions.[4] For more on the clinical literature spanning 650 articles, please see the bibliography.

Step 3: Estimating effect size and scope and evidence for sex differences. We reviewed the clinical literature for each intervention in each of the 64 diseases to identify its effectiveness in mortality and disability reduction in both sexes and in each sex individually where reported.[5] Where

[3] For example, for ischemic heart disease and ischemic hemorrhagic stroke, we identified an intervention that included a polypill (containing antihypertensive and cholesterol-lowering medicines) and lifestyle risk reduction education that incorporated guidance on smoking, substance use, diet, and physical activity.

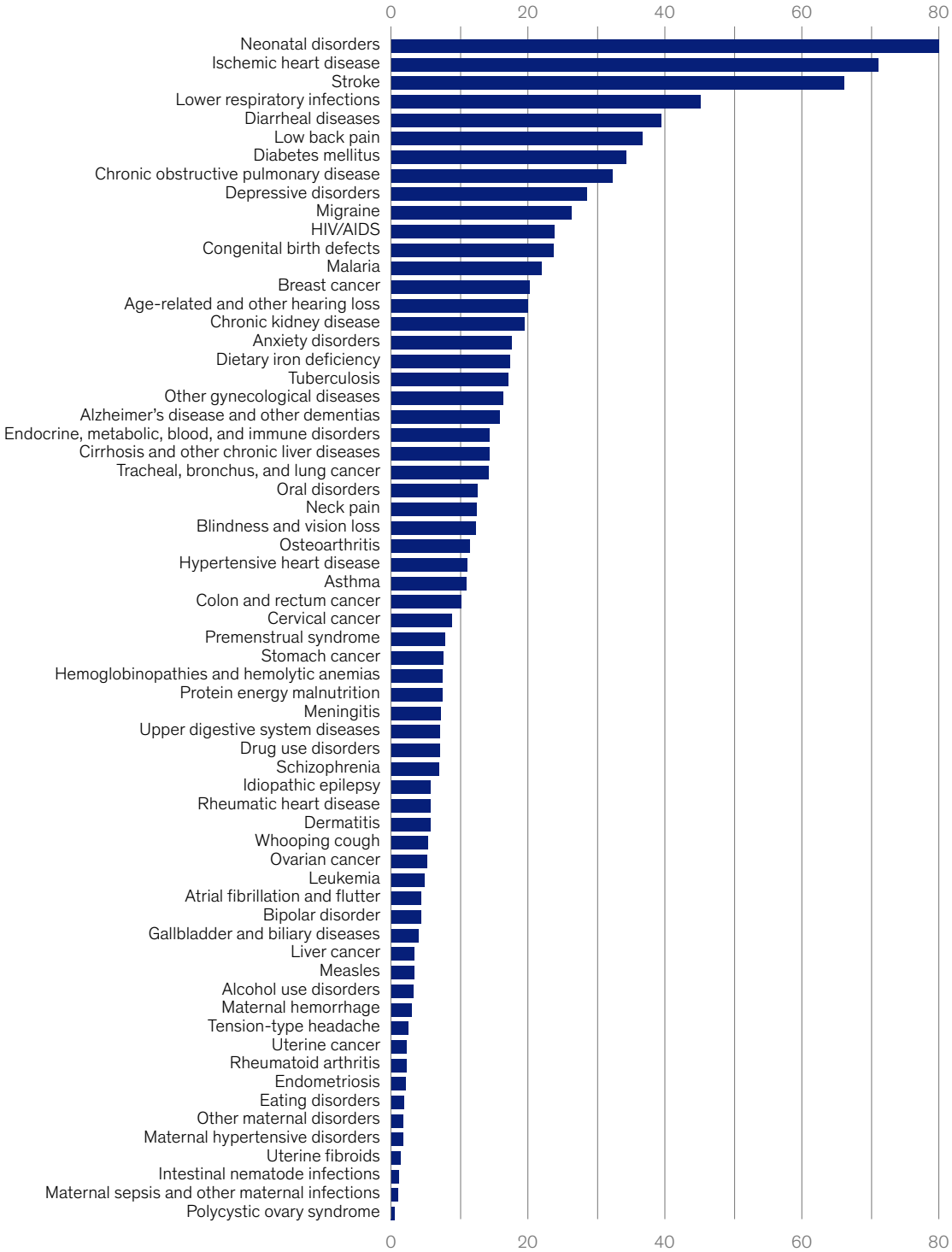
[4] The analysis recognizes that many health interventions are a combination of actions. For instance, a diabetes prevention program includes assessment, individualized nutritional guidance and support, group physical activity sessions and advice, and goal setting and monitoring.

[5] Effectiveness (effect size, efficacy) here refers to the impact of an intervention as measured and reported in a clinical study or trial. The context of a clinical study or trial usually represents "ideal" conditions, which may or may not be attainable in the real world.

Exhibit 1

Sixty-four diseases contribute 86 percent of healthy life years lost to poor health.

Female disease burden of diseases covered in our research, 2019, millions of DALYs[1]



Note: Total = 1.1 billion DALYs global disease burden for females (100% disease burden).
[1]Disability-adjusted life years.
Source: : Institute for Health Metrics and Evaluation (IHME), used with permission, all rights reserved; McKinsey Health Institute analysis

McKinsey & Company

clinical studies used outcome measures in relation to reduction in symptom severity, this was used as a proxy for reduction in the disability burden. Where possible, we referred to papers cited in the international clinical guidelines (used to identify the relevant basket of interventions). This was supplemented with additional literature reviews where necessary. All evidence was categorized using a standardized grading system, where evidence from multiple, high-quality randomized controlled trials was given a higher rating than evidence from less robustly designed studies.[6]

Out of 183 interventions we reviewed, we found effectiveness and/or adoption (uptake or usage in eligible patients) reported for each sex individually (rather than both sexes combined) at the level of statistical significance for 50 percent of the interventions.[7] For 13 percent of the interventions, no sex difference in effectiveness was found. For 32 percent of the interventions, the sex difference in effectiveness was to the advantage of men, and for 5 percent of the interventions, the difference advantaged women.

For interventions where sex-disaggregated effectiveness rates were not reported, we assumed the same proportional effectiveness difference as for the interventions where sex differences had been studied and reported. For interventions used to treat female-specific disorders, we compared the average difference in effectiveness for all interventions in this group with a comparable range of interventions used to treat male-specific disorders.

We then identified the share of the disease burden to which the intervention would apply. In most cases, we assumed the same level of effectiveness across the disease in all age groups. In some cases, specific interventions were shown to be effective in specific age groups or in people diagnosed with specific strains or subtypes of disease. In these cases—for example, where an intervention such as the BCG vaccine for tuberculosis was suitable for use only in children—we assigned an impact only in appropriate age groups.

Step 4: Estimating potential for higher adoption, including evidence for sex differences. We assessed the scope for higher levels of intervention adoption and time effects separately in each of the four country income archetypes,[8] using the following criteria. Peak or best-practice level of adoption that could be achieved was based on aspirational yet realistic assumptions about potential adoption with best-practice levels of delivery and uptake. Wherever evidence was available, we used estimates from published studies or routinely reported statistics to assess current adoption and compared this with best practice or recommended levels. Where sex differences in current adoption were reported, this was included in the analysis and used to calculate an average sex-difference proportion, which was applied to interventions where no sex-disaggregated data was available. Where no adoption estimate (for either sex) could be found in the literature, an expert group agreed on guidelines to determine adoption thresholds for different intervention categories and income archetypes, and these were tested for relevance for each individual intervention and disease with relevant clinical experts.

The health gap for women due to lower average adoption was calculated following the same approach used to calculate the health gap due to sex difference in effectiveness, described in the previous pages.

Step 5: Estimating time lag to impact. We estimated the time required to reach peak adoption rates for different intervention types and, separately, for the four country income archetypes.[9]

[6] For the grading approach, we drew on Gordon H. Guyatt et al., "GRADE: An emerging consensus on rating quality of evidence and strength of recommendations," *BMJ*, April 2008, Volume 336, Number 7650.

[7] This analysis excluded interventions aimed at sex-specific conditions and interventions used in the management of mental health and substance use disorders.

[8] This analysis used the World Bank classification system, which groups countries into four categories based on GNI per capita: low income, lower-middle income, upper-middle income, and high income. Afghanistan and Ethiopia are examples of low-income countries, while India and Kenya are examples of lower-middle-income countries. China and Brazil are the largest upper-middle-income countries, and the United States, Japan, and all countries in Western Europe are examples of high-income countries.

[9] The analysis assumes longer times to scale in low- and middle-income countries based on USAID findings. *Idea to impact: A guide to introduction and scale of global health innovations*, US Agency for International Development, 2016.

Assumptions are based on real-world examples of time to implementation of different types of intervention in different health system contexts, as well as trends in expansion of universal health coverage. Time delay to disease burden impact, or the lag between treatment and effect, varies by disease and is based on an assessment of disease progression. We did not assume any sex differences in time lag to implementation or impact.

Step 6: Sequencing interventions. All interventions were ordered within each disease category to which they applied. This sequencing was based on the principle that primary prevention would be applied before secondary prevention or treatment of established disease. The purpose of the sequencing was to ensure that impact was only applied to a remaining disease burden and that no intervention impact was double counted.

Step 7: Expert review. To test the findings of our detailed analyses, clinical experts in each of the disease areas reviewed the evidence used and assumptions made. They reviewed the basket of interventions identified for each disease, the scope for additional uptake and size of potential health benefits, the intervention sequencing, the time to implementation and impact, and the overall health impact at 2040 for different country income groups.

Extrapolating impact to other health conditions
Approximately 115 level three conditions, which are responsible for just over 14 percent of the global female disease burden, were not covered by the detailed disease review. These conditions were organized into disease groupings based on the IHME GBD 2019 level two hierarchy.

We assumed the health gap due to sex differences in effectiveness or adoption would be consistent with the weighted average for the level two group, based on the detailed analysis of the higher-burden diseases in that group.

Assessing the impact of innovations

To identify the most promising technologies to further reduce the gender health gap by 2040, we focused on technologies with potential impact on diseases with the greatest remaining unmet need and assessed current biological understanding of the disease and the effort and excitement surrounding each, measured by research funding. To estimate the potential impact of innovations, we focused on initiatives that are already in early stages of development or being piloted at a small scale. Efforts in this visible pipeline are more likely to be approved and adopted broadly enough to make a material impact on health over 20 years. We focused on innovations that have a direct impact on health outcomes, such as drugs, digital therapies, devices, and medical technologies.

This is not an exhaustive list but includes the most promising innovations that are in the pipeline or have been piloted at a small scale today. We recognize that there is a certain level of overlap between these categories. The impact of future innovations is inherently uncertain. While our estimates reflect our best assessment of the potential at the time of the report writing, they should not be considered a forecast of 2040 outcomes. A comprehensive review of the methodological approach followed to identify innovations and estimate their future impact is provided in an appendix to the McKinsey Global Institute report *Prioritizing health: A prescription for prosperity*.[10]

We interviewed experts in life sciences research and development to reach a perspective on the likelihood that the technology areas identified would reduce, perpetuate, or exacerbate sex differences in effectiveness and adoption. The consensus from the experts interviewed was that, without radical reform, it is likely that sex differences will be perpetuated at their current levels. A fuller discussion is provided in Chapters 3 and 4 of the report. In line with the expert guidance received, we assumed that the breakthrough technologies would, on average, recreate the same proportional sex differences in effectiveness and adoption as the existing interventions reviewed in detail. Therefore, we applied the same assumptions and process used for the existing interventions where sex differences had not been reported (described in the previous section).

[10] "Prioritizing Health: A prescription for prosperity," technical appendix, McKinsey Global Institute, July 8, 2020.

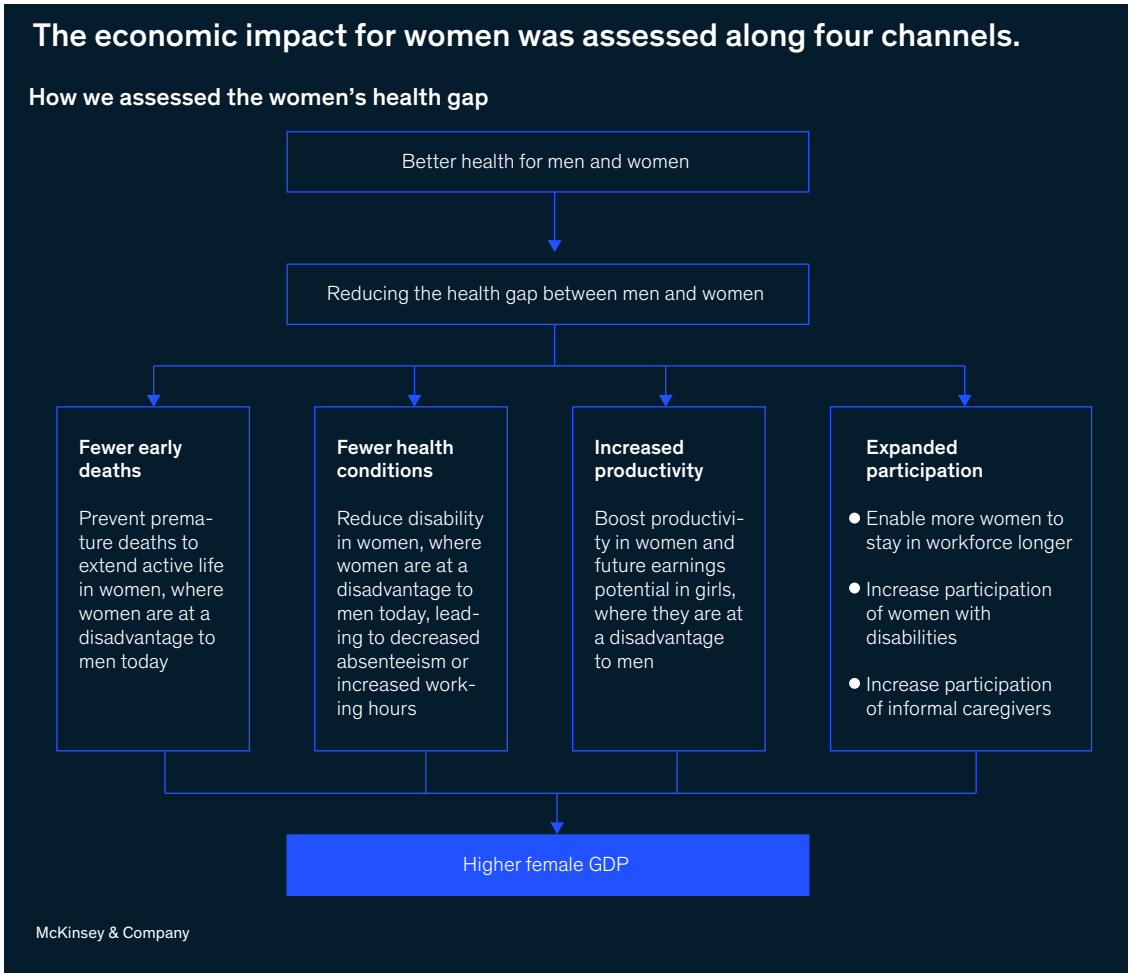
Calculating the economic impact of reducing the women's health gap

To size the economic impact from closing the gender health gap, we estimated the supply side benefits from having a larger, healthier, and more productive female labor force, and we used this to project the annual potential GDP contribution to 2040.[11] Further, to indicatively size the feasibility of reducing the health gap, we estimated the incremental healthcare costs associated with implementing the interventions included in our analysis.

Quantifying the economic benefits from a healthier female labor force

To quantify the potential economic benefits from a healthier population, we assessed the GDP impact along four channels: fewer early deaths, fewer health conditions, expanded labor force participation—all of which expand the number of workers—and increased productivity of healthier worker, which increases the economic contribution per worker (Exhibit 2). Analysis was conducted at the country level and aggregated to produce regional and global estimates.

Exhibit 2



[11] A large share of the economic benefits are realized after 2040 because preventive actions early in life yield benefits decades later.

Fewer early deaths. Improved health due to reducing the women's health gap expands the labor force by reducing premature deaths. For each country, we calculated the impact of mortality reduction for each year from 2020 to 2040, using IHME life expectancy projections, based on the scale of mortality reduction (by disease and age group) calculated in our health gap model.[12] We then translated this health impact into an increase in labor supply. When a premature death is averted by preventing a disease (for example, using antihypertensive and cholesterol-lowering drugs to prevent a heart attack), we assumed affected individuals will have the average labor force participation rate of others in their age group in their country. We relied on the International Labour Organization's historic and projected sex-specific labor force participation rates for each five-year age group, country, and year.[13]

For therapeutic interventions that reduce premature mortality for established health conditions, we adjusted the labor force participation rate with a factor that reflects the lower likelihood of returning to the workforce after a period of ill health. We did this by assigning both a probability of post-treatment return to work (factor between 0 and 1) and a time lag between therapeutic treatment and full return to the workforce, based on labor force participation rate for age group and clinical experts' assessments of return probability.[14] Last, we adjusted the forecast size of the labor force by the unemployment rate because not everyone willing to work may find employment.[15]

To determine the potential GDP gain from expanding the labor supply by averting deaths, we estimated the additional economic output created when a person continues working or returns to work. For averting deaths through prevention (for example, smoking cessation), we multiplied the number of people added to each country's labor supply by the GDP per employed person.[16] For deaths averted through better treatment or management of chronic health conditions, we applied a 5 percent productivity discount for people in high-income countries who were formerly chronically ill, because the evidence suggests that those conditions are negatively correlated with wages.[17]

Fewer health conditions. Improved health for women raises labor force participation by reducing short and long-term absenteeism due to disability. We calculated the reduction in years lived in disability from the disease burden model. We assessed the economic impact separately for health conditions avoided because of preventive interventions and for health conditions improved because of better treatment. For diseases prevented, we applied labor market participation rates, unemployment rates, and productivity levels similarly to the case of early deaths averted. For established diseases treated, we adjusted these estimates with reduced likelihood to reenter the labor force after therapy, as well as lower productivity to reflect lower average education among people with health conditions—again consistent with the approach used for premature deaths averted by treatment of health conditions, described.

[12] "Global Burden of Disease Study 2019 (GBD 2019)," Institute for Health Metrics and Evaluation (IHME), Global Burden of Disease Collaboration Network, 2020.

[13] International Labour Organization Department of Statistics. The 2030 forecasts are the latest available, and the analysis assumed the forecast 2030 rates will apply from 2031 through 2040.

[14] These factors were derived from academic research and interviews with experts. The factors vary widely by disease. For some diseases (such as liver, stomach, and lung cancers), we assumed it is very unlikely that the person will go back to work after treatment and applied a factor of 0 percent. For people diagnosed with less severe diseases, we assumed there would be a higher probability of returning to the labor force after treatment (for example, 25 percent for ischemic stroke, 50 percent for alcohol use disorders). Finally, for diseases where effective treatments are available that allow people to participate fully or almost fully after the treatment, we assumed higher return-to-work probabilities—for example, 80 percent for breast cancer and 100 percent for vision impairment treated with cataract surgery. In each case, the return to work is a proportion of the labor force participation rate for the gender, age group, country, and year. For a literature review of evidence relating to return-to-work probabilities for different conditions, see Daniel Prinz et al., *Health and economic activity over the lifecycle: Literature review*, NBER working paper number 24865, July 2018.

[15] The forecast for unemployment level in 2040 used World Bank unemployment data for 2019.

[16] GDP per employed person from Oxford Economics.

[17] Lower educational attainment and lower wages are associated with chronic diseases. The analysis looked at diabetes and cardiovascular diseases in the United States and Germany, comparing the disease burden with the level of educational attainment, and level of education with level of earnings. See Christin Heidemann et al., "Social inequality and diabetes mellitus—developments over time among the adult population in Germany," *Journal of Health Monitoring*, June 2019, Volume 4, Number 2; Andy I. Choi et al., "Association of educational attainment with chronic disease and mortality: The Kidney Early Evaluation Program (KEEP)," *American Journal of Kidney Diseases*, August 2011, Volume 58, Number 2.

Expanded labor force participation. We considered the potential impact from expanded labor force participation for two groups: informal caregivers, who would not face the same demand to care for family members if overall population health was improved, and people with disabilities who could benefit from better opportunities for economic participation.

— Informal caregivers. With healthier populations, there would be lower demand for informal caregivers, some of whom could choose to undertake paid work instead. The starting point for our analysis was the share of informal caregivers in the populations of Organisation of Economic Co-operation and Development (OECD) countries. We assumed this share could be lowered in proportion to the annual disease burden averted according to our disease burden model. We then calculated the disproportionate representation of women in informal caregiving (using an estimate that women make up 70 percent of informal caregivers on average).[18] Finally, we applied the country's female labor force participation rate to estimate the volume of additional women participating in labor force, applying a ramp-up time of 15 years to full impact. Only the impact of the disproportionate representation of women in the informal caregiver workforce was included in the estimate attributed to the women's health gap.

— People with disabilities. In most countries, people with disabilities are currently not working at the rates they could or would like to, and changing their opportunities is another lever for improving economic output. Disability is an umbrella term encompassing people living with impairments, activity limitations, and participation restrictions.[19] Global prevalence of disability for women is estimated to be about 20 percent of the female population.[20] Given that share of people with a disability rises with age, the prevalence among the working-age population (ages 15 to 64) is likely to be lower, and we assumed a prevalence rate of 13 percent for this age group.[21] For this share in our baseline projection, we relied on labor force participation rates of people with disabilities for 39 countries for 2010, the latest data available, and compared them with the labor force participation rate of the working-age population in the same year.[22]

To estimate the potential increase in the labor force participation rate, we determined a best-practice level of participation (relative to the labor force participation rate in the general population), based on the upper-quartile participation rate for people with a disability among 39 countries. This was roughly 25 percent lower than among the general population.[23] We assumed that with the right incentives and accommodations, people with

[18] Evidence from all age groups indicates that women make up around 60 percent of informal caregivers. However, following discussion with experts, we assumed that women are more likely to play informal caregiver roles at younger ages than men, so we elected to use a higher estimate (of 70 percent) to estimate the proportion in working ages, which is more relevant to our economic analysis. See also Yuke-Lin L. Kong et al., "Factors associated with informal caregiving and its effects on health, work, and social activities of adult informal caregivers in Malaysia: Findings from the National Health and Morbidity Survey 2019," *BMC Public Health*, June 2021, Volume 21; Gaëlle Ferrant, Luca Maria Pesando, and Keiko Nowacka, *Unpaid care work: The missing link in the analysis of gender gaps in labour outcomes*, OECD Development Centre, December 2014; Nandita Bhan, Namratha Rao, and Anita Raj, "Gender differences in the associations between informal caregiving and wellbeing in low- and middle-income countries," *Journal of Women's Health*, October 2020, Volume 29, Number 10; Shyhrete Rexhaj et al., "Women involvement in the informal caregiving field: A perspective review," *Frontiers in Psychiatry*, January 2023, Volume 14; "Women shoulder the responsibility of 'unpaid work,'" UK Office for National Statistics, November 10, 2016.

[19] Disability Fact Sheet, WHO, March 7, 2023.

[20] *Good practices of accessible urban development*, UN Department of Economic and Social Affairs, October 21, 2016; Sophie Browne, *Making the SDGs count for women and girls with disabilities*, issue brief, UN Women, 2017; *Disability, work and inclusion: Mainstreaming in all policies and practices*, OECD, 2022.

[21] L. Kraus et al., *Disability statistics annual report*, Institute of Disability, University of New Hampshire, 2017.

[22] Eurostat, Employment and Unemployment (Labour Force Survey), European Union; International Labour Organisation (ILO) and Organization for Economic Co-operation and Development (OECD), *Labour market inclusion of people with disabilities*, first meeting of the G20 Employment Working Group, Buenos Aires, Argentina, February 20–22, 2018.

[23] Based on OECD data; 75 percent represents the top quartile from across income settings for which data are available. The countries with the highest employment rates for people with disabilities as a share of employment rate for people without disabilities were identified as Switzerland (81 percent), Mexico (79 percent), Canada, France, and Korea (each at 75 percent); *Transforming disability into ability: Policies to promote work and income security for disabled people*, OECD, February 2003.

disabilities could reach 75 percent of the general labor force participation rate of the working-age population in each country by 2040.[24] Where a country was already at this level, we modeled no further uplift. Additional labor force participation was then multiplied by the share of the population with a disability of any kind. We assumed disability prevalence was similar across all countries (because no better comparable estimate is available), and to be conservative, we assumed that the accommodations and perceptions required—for example, disability-inclusive public transport and design of the built environment and policies to address discrimination in recruitment and the wider workplace—would be more likely to occur in urban areas and in larger organizations. To account for this, we assumed a relatively low adoption rate of 20 percent globally and a ramp-up time of 15 years to see the full benefit. The share of this impact attributed to the women's health gap was calculated as the difference between male GDP impact and female GDP impact due to sex-specific differences in disability prevalence.

For all cases of expanded labor force participation, we calculated economic impact as increased labor supply multiplied by GDP per employed person, discounted by the country unemployment rate.

Increased productivity of healthier workers. Healthier women are likely to be more productive. We sized two channels through which improved health can boost productivity: healthier adults who are more productive and experience lower levels of workplace presenteeism, and healthier children who grow up to be more productive adults with higher earnings potential.

— Increased productivity of healthier adults. We reviewed the literature to identify the health conditions that affect the productivity of working adults. The conditions for which better management would have the most impact on productivity, as well as boosting quality of life, included iron deficiency, depressive or anxiety disorders, low-back pain, migraine, endometriosis, premenstrual syndrome, rheumatoid arthritis, osteoarthritis, and menopause.[25] We started by identifying the prevalence of the conditions in each age group, country, and year and then identified the potential reduction in the women's health gap with interventions tackling the diseases, based on the disease reduction model described earlier in this appendix.

  We then assessed the potential productivity gains achievable, based on research into the productivity cost of each disease, which shows that an increase in productivity of as much as 5 percent is possible.[26] To assess the productivity impact, we multiplied the affected population by the labor force participation rate for the specific age group, the share of employed people, the assumed productivity increase after averting the specific disease, and GDP per employed person.

— Increased productivity from healthier childhoods. We looked at three areas where evidence shows that poor health in childhood

[24] For countries for which no data were available, labor force participation uplift was calculated by subtracting the average baseline labor force participation rate of people with disabilities from 75 percent of the average labor force participation rate in the working-age population of countries with available data.

[25] Mark E. Schoep et al., "Productivity loss due to menstruation-related symptoms: A nationwide cross-sectional survey among 32 748 women," *BMJ Open*, June 2019, Volume 9; Ahmed M. Soliman et al., "The effect of endometriosis symptoms on absenteeism and presenteeism in the workplace and at home," *Journal of Managed Care & Specialty Pharmacy*, July 2017, Volume 23, Number 7; Jennifer Whiteley et al., The impact of menopausal symptoms on quality of life, productivity, and economic outcomes," *Journal of Women's Health*, November 2013, Volume 22, Number 11; D. Kim et al., "Importance of obtaining remission for work productivity and activity of patients with rheumatoid arthritis," *Journal of Rheumatology*, 2017, Volume 44, Number 8; Marco dacosta Dibonaventura et al., "Evaluating the health and economic impact of osteoarthritis pain in the workforce: Results from the National Health and Wellness Survey," *BMC Musculoskeletal Disorders*, April 2011, Volume 12; Ronald C. Kessler et al., "Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers," *American Journal of Psychiatry*, 2006, Volume 163, Number 9; Why invest in nutrition?," in *Repositioning nutrition as central to development: A strategy for large-scale action*, World Bank, 2006; James P. Smith and Gillian C. Smith, "Long-term economic costs of psychological problems during childhood," *Social Science & Medicine*, 2010, Volume 71, Number 1; *Copenhagen Consensus 2008 challenge paper: Hunger and malnutrition*, Copenhagen Consensus Center, May 2008; Dan Chisholm et al., "Scaling-up treatment of depression and anxiety: A global return on investment analysis," *Lancet Psychiatry*, May 2016, Volume 3, Number 5.

[26] Dan Chisholm et al., *Lancet Psychiatry*, May 2016; Donna Allen et al., "Four-year review of presenteeism data among employees of a large United States health care system: A retrospective prevalence study," *Human Resources for Health*, November 2018, Volume 16, Number 1; "Why invest in nutrition?," in *Repositioning nutrition as central to development: A strategy for large-scale action*, World Bank, 2006.

affects future productivity: malnutrition before the age of five, substance use disorders, and depressive and anxiety disorders.[27] To calculate the productivity impact, we started with estimates of gender gap reduction potential from the model discussed earlier (including adoption and time lags) and applied them to the estimated age-specific incidence for each condition for each year modeled (2020–40). This enabled us to calculate the number of people benefiting from the interventions in the healthy growth scenario and account for future productivity impact as they join the labor force (estimated using local, age-specific participation and employment rates). The size of the benefit was estimated based on studies of enhanced earnings. The benefits were then multiplied by the GDP per employed person. Given that the health interventions in early life improve lifetime health and earnings potential, the 2040 estimate included in our sizing reflects only a small share of the benefits, as the beneficiaries are relatively young and will continue to be in their prime active years far beyond 2040.

Estimating the cost of narrowing the women's health gap

The feasibility of investing to narrow the women's health gap depends on the affordability of implementing the interventions identified. The cost of healthcare is of particular concern both in aging developed economies that are facing rising healthcare costs and in low-income countries where lack of resources is a major constraint on healthcare service provision. To investigate this question, we compared the cost of delivering the intervention with the value that would be created, in terms of economic return to society.

Incremental costs of implementing intervention

Our health improvement impact modeling is based on measuring disease burden in DALYs. The "cost per DALY averted" metric is one of the most widely available measures of net incremental costs of interventions that can be applied to DALY data. This metric reflects a net unit cost that takes into account both the costs of delivering the intervention to the target population and the savings in treatment costs that are avoided as a result. We identified and prioritized a set of gold-standard sources from which to collect data on cost per DALY averted for the interventions identified in the detailed disease reviews for each of the four income archetypes.[28] The sources were prioritized as follows:

1. The WHO's *Global action plan for the prevention and control of noncommunicable diseases, 2013–2020*, Appendix 3. This source provides estimates that are methodologically consistent across all income groups.

2. *Disease Control Priorities*, third edition (DCP3). This source provides estimates that are methodologically consistent across low- and middle-income settings.

3. Tufts Medical Center Health Economics Database. This source is methodologically sound but potentially inconsistent between individual estimates because it is a collection of reviewed papers, rather than a standardized approach.

We relied on a limited set of established sources, rather than a more exhaustive study of the primary literature, to maximize methodological consistency and comparability with other external work for a metric that can be measured in diverse ways. However, we recognize that the exact methodology varies between sources and includes many complex variables that could differ between and within countries, such as price levels of products and supplies, salary levels of healthcare workers, and societal costs of informal caregiving or lost productivity. The cost analysis should be interpreted

[27] In most studies, productivity is measured as earnings potential. Daniel Prinz et al., *Health and economic activity over the lifecycle: Literature review*, NBER working paper number 24865, July 2018; Smith and Smith, "Long-term economic costs," July 2010; Sue Horton, Harold Alderman, and Juan A. Rivera, *Copenhagen Consensus 2008 challenge paper: Hunger and malnutrition*, Copenhagen Consensus Center, May 2008.

[28] We reviewed other sources, including WHO's OneHealth and IHME's Healthdata, but did not include them because they provided limited incremental value and increased methodological complexity.

as directionally indicative and not a precise forecast of actual costs facing any country or health institution.

For more than two-thirds of the interventions assessed, we identified a cost per DALY averted value from one of our three prioritized sources. We were able to identify a strong analogy for around one-quarter of the interventions. For example, for clean peri- and postnatal practice, we used the DCP3 value for intrapartum care, which includes clean practice. For around 10 percent of all the cases, we found only weak analogies. For example, for oral therapies used in migraine prevention, we took the cost per DALY averted value for episodic treatment with newer antidepressant therapy from DCP3 because several antidepressants are used as second- and third-line treatments for migraine, and treatment can be considered episodic and primary.

Where cost per DALY averted estimates were reported as a range rather than a precise figure, we used the midpoint of the range for all income archetypes. Where a range was provided only for upper-middle- and high-income countries, the high point of the range was used for high-income countries. Health economics experts reviewed this approach. Where only an upper range was available (for example, more than $15,000), we continued to review the priority sources to identify a more precise estimate.

To estimate differences in cost by income archetype, we used the following methodology. Estimates of cost per DALY averted were collected for all interventions and all income archetypes from the three gold-standard sources (searched in order of priority), creating a core data set with some missing values where estimates could not be found. This core data set was used to calculate weighted average ratios between income archetypes: high income = 1, upper-middle income = 0.6, lower-middle income = 0.3, low income = 0.4. These ratios were used to derive cost per DALY averted values for selected interventions where estimates had been identified in the literature for one or more income archetypes but not for all four.[29] Where interventions involved multiple components for which estimates of cost per DALY averted were available only for individual components, we selected the highest-cost component and used this as the estimate.

All estimates were converted from the reporting currency into a standard currency of constant dollars, using World Bank data. Costs were multiplied by the volume of DALYs averted (in the given country and year) to calculate the incremental costs of delivering the interventions. The costs of interventions for conditions not analyzed in detail in the disease review, and therefore not included in the cost per DALY averted data collection exercise, were estimated based on the weighted average cost per DALY averted (by income archetype) for all the interventions individually analyzed in the detailed disease reviews.

Cost-effectiveness of interventions

Realistically, only cost-effective interventions would be implemented. Therefore, in sizing total costs of interventions, only interventions that would be considered cost-effective within the income archetype were included in the analysis.[30]

Total costs were estimated at intervention and disease level by multiplying cost per DALY averted by the DALYs averted for each country and year. These values were compared with the GDP contribution for the same intervention, disease, country, and year to estimate the economic return at income archetype level.

[29] For example, one of the core sources used was the updated Appendix 3 of *Global action plan for the prevention and control of noncommunicable diseases, 2013–2020*, WHO, April 2017. It provides estimates of cost per DALY averted for two aggregated income archetypes: a combined category for low- and lower-middle-income countries and a combined category for upper-middle- and high-income countries.

[30] Cost-effectiveness was defined as a cost per DALY averted not higher than three times GDP per capita for each country income archetype.

Designed by McKinsey Global Publishing