Online Appendix 2. Detailed description of methods for developing the core basket of SDH action indicators

Overview of the systematic identification and prioritization of measurement themes, concepts, and core basket of indicators

An expert Working Group (WG) process was used to create a basket of 15-25 core indicators with strong content validity for measurement concepts drawn from the themes of the Rio Political Declaration (see Table 1 for glossary definitions of these terms). To develop this basket of core indicators, we followed a standard process of systematic indicator identification, assessment, and prioritisation. First, we identified and systematically prioritized key measurement themes (see Online Appendix 3 for definitions of measurement themes) related to action on the SDH in the Rio Declaration and then developed key measurement concepts to linking the themes to the Rio Declaration pledges (see Online Appendix 4 for the rationales behind linking pledges to Measurement concepts and see Appendix 1, Table A1.1 for the criteria for selecting the measurement concepts).

Next we systematically identified and selected indicators that met inclusion and quality criteria (Table 4) for each measurement concept. We identified an initial set of indicators contained in formal databases, monitoring systems including the 230 preliminary SDG indicators, or monitoring reports. The potential candidate indicators meeting inclusion/exclusion criteria (see Appendix 2, Table A2.1 for inclusion/exclusion criteria for the long list of candidate indicators) were compiled into a Long Indicator List. Throughout this process the measurement concepts were also revisited and revised to reflect WG priorities and also to better align with the available indicators. To ensure alignment with the 2030 Sustainable Development Agenda,[13] suitable SDG indicators that measure relevant action on the SDH were used as the foundation, but were replaced if better indicators from other sources were identified. We systematically assessed each candidate indicator in the Long List by several quality criteria (Table 4 for quality criteria). The prioritized indicators were then compiled into the Short Indicator List: the proposed basket of core indicators. WHO regional focal points and two member states were consulted preliminarily for feedback on the measurement quality and suitability of the selected indicators.

Identification of measurement themes

The WG reviewed each pledge of the Rio Declaration to identify the key measurement themes communicated in each pledge. The frequency of appearance of each theme across the pledges in all action areas (referred to as 'domains' in the monitoring system) was first tallied to guide identification of salient measurement themes. Themes covered policy sectoral entry points communicated by pledges such as social service and protection policies or development strategies/policies/rights. For pledges that also referred to directly or indirectly to disadvantaged populations, such as Indigenous People, children, women, transgender people, and informal workers, these core populations were also identified as part of the measurement themes. Over several WG meetings, members highlighted pledges and further prioritized the themes that best reflected the domains. Measurement themes were neither unique to individual domains, nor specific to pledges. External experts were consulted on the prioritization of the measurement themes. This included feedback from experts from each of the six WHO Regional Offices, as well as from key academics. Measurement themes were too broad to represent specific action areas for monitoring, therefore they were coupled with ideas contained in

pledges to generate measurement concepts. For a glossary of the terms used (e.g. measurement theme, concept, pledge, action area, etc.) please see Table 1. A definition for each measurement theme is provided in Online Appendix 3.

Identification and prioritization of measurement concepts

Measurement themes were too broad to represent specific domains for monitoring. Therefore, in WG meetings and in one-on-one interviews with WG members, the broader measurement themes (acting as a foundation for framework development) were linked with foundational ideas in the pledges to generate and refine measurement concepts (See Online Appendix 2 for rationale linking themes to measurement concepts). WG members then developed and applied criteria for selecting priority measurement concepts (see Appendix 2, Table A2.2 for list of criteria to select measurement concepts). Subsequently the WG systematically applied these criteria and jointly selected the core measurement concepts. The core measurement concepts were updated throughout the expert WG process in order to best reflect the measurement priorities (e.g. equity and vulnerable populations) and the availability of high quality indicators.

Identification and prioritization of indicators

First, WG members and consulted external experts proposed 21 sources for candidate indicators (Online Appendix). The secretariat screened these databases, reports, and other indicator sets including the 230 SDG indicators presented in April 2016 [14] for relevant SDH action indicators. In order to determine a strong set of criteria for evaluating indicators for action on the SDH, we reviewed existing relevant reports of international monitoring efforts as well as the indicator development literature to ensure consistency with best practices of indicator selection. The WG collaboratively developed a tailored list of inclusion/exclusion criteria for screening indicators (Appendix 2, Table A2.1) as well as quality assessment criteria (Appendix 2, Table A2.2) to select the best and most relevant indicators for measuring action on the SDH.

SDH action indicators that fulfilled the Inclusion Criteria listed in Appendix 2, Table A2.1 were compiled into the Long Indicator List. Indicators aligned with the SDGs were prioritized for inclusion in the basket of core indicators. For each measurement concept, a list of process-level SDGs and indicators from the Long List were consulted, and relevant indicators were chosen for each measurement concept when available.[18] In the event that no existing indicators were identified for a given measurement concept, gaps in measurement were underscored. Each identified indicator was assessed based on (1) its face-validity (or match) with its corresponding measurement concept and (2) its technical quality. First, each indicator was scored on measurement concept match criteria (/7). A score of 1 indicated that the indicator was not at all in accordance with the corresponding measurement concept, while a score of 7 indicated that it was completely in accordance with the corresponding measurement concept; a rationale for the score was provided.

A second activity involved WG members scanning SDG indicators and other indicators on the Long Indicator List for suitability to each measurement concept. Select WG members numerically rated candidate indicators according to the quality assessment criteria listed in Table 7. Based on the number of quality assessment criteria met, the indicator was given a technical quality rating (/8). A rationale for the score was provided, and data availability (i.e. current, long-term, or no availability) is provided in Online Appendix 6.