Institute for Global Prosperity

Measuring What Matters:

A Systematic Review of Social Service Coverage in Beyond-GDP Indicators

Nikolaos Tzivanakis, Georgios Melios, Henrietta L. Moore

Working Paper

January 2026

About the Institute for Global Prosperity, UCL

The Institute for Global Prosperity at UCL (IGP) is redesigning prosperity for the 21st century, changing the way we conceive and run our economies, and reworking our relationship with the planet. IGP's vision is to build a prosperous, sustainable global future, underpinned by the principles of fairness and justice, and allied to a realistic, long-term vision of humanity's place in the world.

The IGP undertakes pioneering research that seeks to dramatically improve the quality of life for current and future generations. Its strength lies in the way it allies intellectual creativity to effective collaboration and policy development. Of particular importance to the IGP's approach is the way in which it integrates non-academic expertise into its knowledge generation by engaging with governments, policy makers, business, civil society, the arts and local communities.

Institute for Global Prosperity

UCL

ISBN: 978-1-913041-64-9

January 2026

Authors:

Nikolaos Tzivanakis[*a], Georgios Melios[b], Henrietta L. Moore[a]

[*] Corresponding author: n.tzivanakis@ucl.ac.uk
[a] Institute for Global Prosperity, University College London
[b] Department of Behavioural Science, London School of Economics and Political Science; Ethos Lab – Centre for Governance and Sustainability Research, Athens, Greece

Contents

Abstract

The Beyond-GDP movement has produced sophisticated frameworks for measuring national progress, yet no systematic assessment has examined whether these frameworks adequately capture social services – the interventions through which welfare states improve citizens' lives. This paper addresses that gap through meta-analysis of 66 Beyond-GDP indicators developed between 1972 and 2023, mapped against a 22-domain taxonomy using keyword-based content analysis with whole-word matching. Our findings reveal a fundamental asymmetry in measurement development. Health (86%), Knowledge & Skills (80%), Material Wellbeing (77%), Work & Job Quality (73%), and Economic Security (70%) demonstrate comprehensive coverage reflecting decades of methodological investment. Care services, by contrast, appear in only 17% of indicators, dropping to just 6% in the full database of 209 indicators. This nearly threefold decrease represents the most severe measurement gap identified, confirming that care for children, persons with disabilities, and elderly individuals remains largely invisible in progress measurement. Service integration (11%) and prevention (6%) prove similarly neglected. Co-occurrence analysis reveals these gaps are structural rather than incidental. Care services achieve below 20% integration with established measurement clusters, compared with above 70% among core economic and social domains. This isolation indicates that expanding existing frameworks will not organically incorporate care; deliberate architectural redesign is required. Methodological limitations compound content gaps: only 12% of frameworks support intersectional analysis, disability disaggregation appears in just 12% of indicators, and publication lags averaging 18 months constrain policy responsiveness. These findings establish that Beyond-GDP measurement, despite substantial conceptual progress, inadequately captures dimensions central to welfare state investment, providing an evidence base for developing more comprehensive social service evaluation frameworks.

Keywords: Beyond-GDP; wellbeing indicators; social services; care economy; meta-analysis; welfare state; measurement frameworks

1. Introduction

For more than half a century, Gross Domestic Product has served as the dominant metric of national progress, shaping political accountability, resource allocation, and public understanding of societal advancement. GDP's appeal is readily apparent: conceptual clarity, quarterly availability, and international comparability have secured its position as the headline indicator around which economic policy revolves. Yet the limitations of GDP as a comprehensive measure of progress have been recognised since the national accounts framework was first established. Simon Kuznets, the architect of modern national income accounting, cautioned as early as 1934 that 'the welfare of a nation can scarcely be inferred from a measure of national income' (Kuznets, 1934, p. 7). His warning proved prescient. GDP captures the monetary value of market transactions whilst remaining systematically blind to distributional equity, environmental sustainability, unpaid domestic labour, health outcomes, educational quality, social cohesion, and the subjective experience of life itself.

These conceptual limitations translate into consequential policy distortions. When GDP growth serves as the primary criterion for evaluating government performance, economic expansion receives policy primacy regardless of how gains are distributed across the population or what costs are imposed on environmental systems and social relationships. Public services that generate substantial improvements in human wellbeing but produce only modest market transactions, care for elderly relatives, support for persons with disabilities, early childhood development programmes, appear peripheral to national progress when assessed through GDP's lens. The metric treats the labour of a care worker supporting a person with dementia to live with dignity as equivalent in value to any other form of market exchange, whilst rendering entirely invisible the far greater quantity of unpaid care work performed predominantly by women within households.

Recognition of GDP's inadequacy has generated sustained efforts over five decades to develop alternative measures capable of capturing the dimensions of human flourishing that market-based accounting systematically neglects. What has come to be known as the Beyond-GDP movement emerged in the 1970s through early quality-of-life indices and pioneering attempts to adjust national accounts for environmental degradation and social costs. The Genuine Progress Indicator developed by Cobb, Halstead and Rowe (1995) and the Index of Sustainable Economic Welfare proposed by Daly and Cobb (1989) represented ambitious efforts to modify GDP by subtracting costs such as pollution and adding benefits such as household labour. The movement gained institutional legitimacy when the United Nations Development Programme launched the Human Development Index in 1990, demonstrating that a simple composite of health, education, and income indicators could provide meaningful cross-country comparison without reducing progress to economic growth alone.

The 2008 global financial crisis and its aftermath proved catalytic for Beyond-GDP measurement. The influential report of the Stiglitz-Sen-Fitoussi Commission (2009), convened by French President Sarkozy, provided both intellectual legitimacy and political momentum for alternative progress measurement, arguing that statistical systems must expand to encompass quality of life, sustainability, and distribution alongside economic production. The subsequent decade witnessed remarkable proliferation: national wellbeing dashboards emerged across OECD member states, with at least 18 countries developing bespoke frameworks tailored to domestic policy contexts. The Sustainable Development Goals adopted in 2015 established a global framework explicitly balancing economic, social, and environmental dimensions across 17 goals and 169 targets. The European Union's Beyond GDP initiative, launched in 2007 and reinforced through the European Pillar of Social Rights, positioned comprehensive wellbeing measurement as essential infrastructure for evidence-informed social policy across the continent.

Yet despite this remarkable expansion in alternative progress measurement, a critical blind spot has persisted largely unexamined. Social services, the

concrete interventions through which welfare states translate policy commitments into improved lives, receive inadequate attention in the very frameworks designed to capture what GDP misses. This gap is both paradoxical and consequential. Healthcare and education feature prominently across Beyond-GDP indicators, reflecting robust international statistical infrastructure developed over decades and widespread recognition of their centrality to human development. Material living standards prove comprehensively measured through indicators of income, consumption, wealth, and economic security. Environmental sustainability has achieved growing integration, particularly regarding climate change, air quality, and natural resource depletion (Kubiszewski et al., 2025).

Care services, however, remain systematically underrepresented despite their essential role in supporting vulnerable populations. The International Labour Organization estimates that 2.1 billion people globally require care, children needing supervision and nurturing, persons with disabilities requiring assistance with daily living, and elderly individuals no longer able to manage independently (ILO, 2018). This care is provided through complex arrangements combining formal services, informal family support, and market provision, yet the frameworks designed to measure national progress beyond GDP largely fail to capture whether these care needs are being adequately met. The invisibility of care in progress measurement perpetuates the broader invisibility of care work in economic accounting. Women perform over three-quarters of unpaid care work globally (UN Women, 2016), and the failure to measure care adequacy in wellbeing frameworks reinforces the devaluation of this feminised labour.

Service integration presents an equally striking measurement gap. Individuals with complex needs – those experiencing homelessness alongside mental health difficulties, older persons requiring coordinated health and social care, families facing multiple disadvantages spanning employment, housing, and child welfare, depend on services working together across traditional sectoral boundaries. Yet the frameworks designed to capture progress beyond economic output overwhelmingly measure health, education, housing, and social protection as separate domains, failing to capture whether services coordinate effectively to address needs that do not respect administrative categories. Prevention and early intervention represent perhaps the most profound measurement absence. Frameworks readily capture crisis responses (hospital admissions, benefit claims, emergency housing placements) whilst remaining largely unable to assess whether proactive strategies successfully avert problems before they require intensive intervention.

This systematic neglect of social service dimensions within Beyond-GDP measurement carries substantial consequences for policy and practice. Resource allocation decisions informed by incomplete wellbeing measures risk perpetuating underinvestment in precisely those dimensions that remain unmeasured. Performance management systems optimising the indicators that happen to be available create perverse incentives, directing attention toward easily quantified outputs whilst neglecting the complex qualitative dimensions (dignity, respect, continuity, coordination) that matter most to service users. Policy evaluation frameworks that assess economic impacts whilst ignoring service experiences provide distorted evidence regarding intervention effectiveness. And democratic accountability suffers when citizens lack transparent information regarding service quality, accessibility, and outcomes that would enable informed judgment of how collective resources are being stewarded.

This paper addresses these gaps through systematic meta-analysis examining how Beyond-GDP indicators developed over five decades treat social services. The analysis pursues three interrelated objectives. First, it assesses the comprehensiveness with which existing frameworks address domains central to social service provision, identifying which dimensions receive attention and which remain systematically neglected. Second, it examines the methodological characteristics of Beyond-GDP measurement, data sources, update frequencies, disaggregation capabilities, aggregation approaches, and their implications for applying these frameworks to service evaluation contexts. Third, it investigates whether empirical patterns of domain co-occurrence reveal

underlying conceptual structures, identifying which dimensions cluster together in existing practice and which remain isolated from mainstream measurement.

The analysis builds on the comprehensive database of 213 wellbeing indicators compiled by Kubiszewski et al. (2025), applying systematic screening criteria to identify frameworks with relevance to social service evaluation. The resulting sample of 66 indicators spans the full history of Beyond-GDP measurement from Bhutan's Gross National Happiness concept introduced in 1972 through to contemporary dashboards and composite indices. By mapping these frameworks against a 22-domain taxonomy and analysing patterns of domain co-occurrence, the paper provides the first systematic assessment of how alternative progress measures relate to social service evaluation, establishing an evidence base for developing more comprehensive approaches.

2. Methods

This meta-analysis employed systematic review methodology adapted from health sciences for application to policy and indicator literature. The research followed PRISMA guidelines for transparent reporting (Moher et al., 2009), though substantial adaptation was required given that the objects of analysis, indicator frameworks rather than intervention studies, differ fundamentally from clinical research. The methodology comprised four sequential phases: comprehensive identification of candidate indicators, systematic screening against eligibility criteria, detailed assessment of selected frameworks, and synthesis of patterns across the included sample.

2.1 Identification and Search Strategy

The identification phase sought to capture the full universe of Beyond-GDP indicators documented in academic and policy literature. Primary database searches encompassed Google Scholar, Web of Science, OECD iLibrary, the UN Statistics Division repository, and Eurostat, conducted between April and August 2025. Search terms combined Beyond-GDP terminology with wellbeing and sustainability concepts, including variations of 'wellbeing indicator', 'prosperity index', 'happiness measure', 'quality of life indicator', 'social progress index', 'sustainable development indicator', and 'beyond GDP framework'. These searches were supplemented by targeted examination of recent outputs from EU Horizon research projects addressing wellbeing measurement, including WISE Horizon, MERGE, and SPES, alongside systematic review of publications from the European Commission's Beyond GDP initiative, national statistical offices across OECD and EU member states, and grey literature from influential think tanks and civil society organisations.

The search strategy employed snowball sampling from reference lists of major reviews in the field, including Hoekstra's (2019) comprehensive assessment of Beyond-GDP measurement and Costanza et al.'s (2014) influential Nature commentary calling for GDP's replacement. Most significantly, the review built upon the systematic database compiled by Kubiszewski et al. (2025), which employed semantic analysis to classify 213 distinct wellbeing indicators identified across academic and policy literature. This database, representing the most comprehensive inventory of Beyond-GDP frameworks assembled to date, served both as a primary source ensuring no major framework was overlooked and as a cross-reference validating that independent searches had achieved adequate coverage.

2.2 Screening and Selection

Recognising that not all Beyond-GDP indicators are equally relevant for social service evaluation, the review employed a multi-stage screening process designed to identify frameworks most applicable to this purpose whilst maintaining sufficient diversity to characterise the broader measurement landscape. Six criteria were applied sequentially, with the first four serving as binary filters establishing minimum thresholds and the final two employing scaled scoring to assess relevance. This approach balanced methodological rigour with pragmatic recognition that social service evaluation represents a specific

application rather than the primary purpose for which most Beyond-GDP frameworks were designed.

The temporal relevance criterion retained indicators either active between 2000 and 2024 or representing foundational pre-2000 frameworks that significantly influenced subsequent measurement practice. This dual threshold ensured the analysis captured contemporary measurement whilst preserving historically influential frameworks such as the Human Development Index and Gross National Happiness that established conceptual foundations still shaping current practice. Application of this criterion to the initial universe of 213 indicators excluded 26 measures, primarily historical quality-of-life indices from the 1970s and 1980s discontinued without lasting methodological influence, yielding 187 potentially relevant frameworks.

The scale of implementation criterion required sufficient geographic scope to demonstrate transferability beyond single local contexts. Three alternative pathways enabled inclusion: documented deployment across at least three countries, indicating proven applicability across diverse national contexts; systematic application within EU or OECD member state statistical systems, demonstrating relevance to advanced welfare state settings; or recognition as an international reference framework by major multilateral organisations, indicating influence on measurement discourse even where implementation remained limited. This criterion excluded 63 indicators comprising single-city measures lacking evidence of replication, institutional indices developed for internal organisational use, and pilot projects without documented adoption beyond initial contexts. The remaining pool comprised 124 indicators with demonstrated scale or influence.

The public documentation criterion required accessible English-language methodology enabling independent assessment. Whilst necessary given research team capacity and the requirement for systematic assessment, this language restriction may underrepresent innovative non-Anglophone frameworks, particularly indigenous approaches such as Ecuador and Bolivia's Buen Vivir (sumak kawsay), which embed fundamentally different conceptualisations of collective wellbeing and human-nature relationships that challenge Western indicator paradigms. Similarly, culturally-specific frameworks from Japan, Thailand, and the Arab world that incorporate Buddhist, Shinto, or Islamic concepts of happiness may be underrepresented due to language barriers and publication practices. These gaps reflect broader challenges in the Beyond-GDP field, where English-language publication and OECD-aligned frameworks dominate international databases. Six indicators were additionally excluded due to proprietary methodologies, incomplete documentation, or frameworks announced but not yet operationalised, yielding 118 adequately documented frameworks.

The multi-dimensional scope criterion mandated coverage of at least three distinct measurement domains to qualify as comprehensive progress frameworks rather than single-issue metrics. The Beyond-GDP movement explicitly seeks multi-dimensional assessment transcending narrow economic or environmental focus, making breadth a defining characteristic. This criterion excluded 29 indicators focused on single thematic areas – pure environmental metrics, isolated demographic indices, narrow sectoral measures – leaving 89 frameworks meeting all four binary criteria.

The remaining frameworks were assessed on two relevance dimensions using four-point scales. Social content prominence evaluated the extent to which social welfare, equity, and inclusion featured in conceptual architecture, with scoring ranging from 3 (social dimensions central with multiple integrated domains) through 0 (purely economic or environmental focus). Service domain relevance assessed specific applicability to social service evaluation, with scoring ranging from 3 (explicit measurement of service delivery, access, quality, or user outcomes) through 0 (no discernible service connection). Final inclusion required combined scores of at least 3 across these dimensions, ensuring selected frameworks demonstrated meaningful social content and service relevance. This threshold excluded 17 indicators with predominantly environmental or economic focus lacking social service applicability, yielding 72 frameworks meeting all criteria.

A final recency verification excluded frameworks discontinued more than seven years before the review date to ensure analysis reflected contemporary measurement practice rather than historical approaches no longer considered viable. Six frameworks were excluded: the Environmental Sustainability Index (replaced 2006), Green GDP China (discontinued 2007), National Accounts of Wellbeing (single application 2009), Where-to-be-born Index (discontinued 2013), Measures of Australia's Progress (discontinued 2014), and Sustainable Society Index (discontinued 2016). The Human Poverty Index, though discontinued in 2010, was retained as direct methodological predecessor to the active Multidimensional Poverty Index, illustrating evolution within the UNDP measurement programme. The final sample comprised 66 indicators spanning five decades of Beyond-GDP development.

2.3 Analytical Framework

The analytical framework combined systematic domain mapping with co-occurrence analysis to characterise both the content and structure of Beyond-GDP measurement. Domain mapping employed a 22-category taxonomy adapted from the MERGE project (Rum et al., 2024), which had identified approximately 20 distinct wellbeing domains through unsupervised semantic clustering of indicator components. Three modifications aligned this taxonomy with social service evaluation objectives: domains with minimal service relevance were excluded (Mineral Resources, Financial Capital); service-specific domains absent from the original classification were added (Care Services, Service Integration, Prevention); and conceptually overlapping categories were consolidated (Knowledge Capital merged into Knowledge and Skills; Trust and Social Connections distinguished from general Social Capital). The resulting taxonomy enabled systematic mapping of each indicator against 22 domains spanning health, education, economic circumstances, environmental conditions, governance, and social relationships (see Appendix C for complete definitions).

Domain co-occurrence analysis examined which dimensions appear together across indicators, calculating pairwise frequencies to identify empirical clustering patterns. High co-occurrence rates indicate domains that existing frameworks treat as conceptually related or mutually constitutive dimensions that measurement developers tend to include together. Low co-occurrence identifies isolated domains that appear sporadically without consistent integration into broader frameworks. This analysis enabled both empirical validation of pillar structures commonly proposed in Beyond-GDP literature and, critically, identification of systematically neglected dimensions requiring deliberate attention in comprehensive frameworks.

2.3.1 Keyword-Based Content Analysis

Domain coverage was assessed through keyword-based content analysis of indicator component descriptions following a transparent and reproducible methodology. For each indicator, all component descriptions from the Kubiszewski et al. (2025) database were concatenated into a single text string and converted to lowercase to enable case-insensitive matching.

The analysis employed whole-word matching using regular expression word boundaries rather than simple substring matching. This methodological choice prevents false positives where keywords appear as substrings within unrelated terms; for example, ensuring 'road' does not match 'broadband', 'cultural' does not match 'agricultural', and 'green' does not match 'greenhouse'. An indicator was classified as covering a domain if at least one keyword from that domain's curated dictionary appeared as a complete word within its concatenated component descriptions.

Keyword dictionaries were developed following three principles. First, domain specificity: keywords were selected to be distinctively associated with the target domain, with problematic terms generating false positives replaced by more specific alternatives. Second, terminological breadth: dictionaries include alternative phrasings, British and American spelling variants (e.g., 'labour'/'labor'), plural forms (e.g., 'school'/'schools'), and both technical and lay terminology to maximise detection of relevant content. Third, iterative refinement: initial dictionaries were tested against sample indicators and refined to

balance sensitivity (detecting true coverage) against specificity (avoiding false positives).

Wellbeing domains are not mutually exclusive, and some keywords legitimately span multiple domains. For instance, 'income' and 'poverty' are relevant to both Material Wellbeing (as measures of living standards) and Economic Security (as indicators of financial resilience). Rather than artificially forcing exclusivity, the methodology permits conceptually justified overlap where terms genuinely address multiple domains.

Complete keyword dictionaries are provided in the Methodology Supplement to enable replication and scrutiny.

2.4 Data Extraction and Quality Assurance

A structured extraction template was developed through pilot testing on ten diverse indicators before full application. For each framework, the template captured basic metadata (name, developer, year introduced, geographic scope, current status), conceptual architecture (stated theoretical foundation, target audience, policy applications), methodological characteristics (construction type, number of components, aggregation approach, weighting methodology, data sources, update frequency, publication lag), and social service relevance (coverage of service domains, measurement approach, disaggregation capabilities across equity dimensions). Extraction relied on primary documentation including technical reports, academic publications, and institutional websites, supplemented by secondary sources where primary documentation proved incomplete.

Quality assurance procedures addressed both screening decisions and domain classification. For screening, two reviewers independently assessed 20% of indicators against inclusion criteria, achieving 91% inter-rater agreement with discrepancies resolved through discussion and reference to original documentation. For domain classification, two reviewers independently validated keyword classifications on a randomly selected sample of 52 indicators (25% of the full database of 209 indicators), manually inspecting component descriptions and comparing independent judgments against automated keyword classifications. This validation procedure achieved 92% agreement, indicating high reliability of the keyword-based approach. External validation cross-referenced the final included set against indicators identified in recent systematic reviews of Beyond-GDP measurement (Hoekstra, 2019; Costanza et al., 2014), confirming that no major documented frameworks had been inadvertently excluded. Approximately 30% of included indicators lacked publicly available complete component inventories, requiring reliance on secondary descriptions that may not fully capture methodological nuance, a limitation acknowledged below.

3. Findings

The 66 indicators meeting inclusion criteria represent a diverse landscape of Beyond-GDP measurement spanning institutional contexts, geographic scales, and methodological approaches. Before examining domain coverage and co-occurrence patterns, it is useful to characterise this landscape to contextualise subsequent findings. The complete indicator inventory appears in Appendix B.

3.1 The Beyond-GDP Measurement Landscape

The temporal distribution of indicator development reveals three distinct waves corresponding to evolving intellectual currents and policy contexts. The first wave, extending from 1972 through 1999, produced seven foundational frameworks representing 11% of the sample. This period established conceptual approaches that continue to influence contemporary practice. Bhutan's Gross National Happiness, introduced in 1972 and subsequently operationalised through comprehensive surveys, demonstrated that alternative progress concepts could anchor national policy. The Human Development Index launched by UNDP in 1990 proved particularly influential, showing that a parsimonious composite of health, education,

and income indicators could provide meaningful international comparison whilst explicitly rejecting GDP's equation of progress with economic growth. The Gender Development Index (1995) extended this approach to distributional concerns. These pioneering frameworks, whilst limited in number, established feasibility and provided templates for subsequent development.

The second wave, spanning 2000 to 2009, marked significant expansion with 16 indicators representing 24% of the sample. This acceleration followed the Millennium Development Goals and reflected growing recognition of GDP's limitations among policymakers and civil society. The Environmental Performance Index (2006) brought environmental dimensions into comparative assessment. The Legatum Prosperity Index (2007) demonstrated think-tank capacity for sophisticated measurement. Multiple national initiatives emerged as countries began developing bespoke frameworks tailored to domestic circumstances. The 2008 financial crisis and its exposure of the disconnect between economic indicators and lived experience provided further impetus, though the full flowering of post-crisis measurement would come in the subsequent decade.

The third and largest wave, from 2010 through 2023, generated 43 indicators representing 65% of the sample. The Stiglitz-Sen-Fitoussi Commission report (2009) proved catalytic, providing intellectual legitimacy and political cover for statistical offices to invest in comprehensive wellbeing measurement. National wellbeing dashboards proliferated across OECD member states, the United Kingdom, New Zealand, Italy, Germany, and others developed frameworks explicitly designed for policy integration. The 2015 Sustainable Development Goals established global architecture whilst simultaneously stimulating national and subnational measurement initiatives. Recent frameworks demonstrate growing sophistication in equity analysis, subjective wellbeing assessment, and integration with policy cycles, suggesting maturation from advocacy tools toward operational performance systems.

The institutional distribution reveals a balanced ecosystem across five categories. International organisations including UN agencies, OECD, World Bank, and EU institutions account for 30% of frameworks (20 indicators), dominating global comparative measurement through established statistical infrastructure and intergovernmental mandates. National governments, primarily OECD members, represent 29% (19 frameworks), developing wellbeing dashboards increasingly integrated with budget processes and parliamentary reporting. Academic institutions contribute 23% (15 frameworks), often introducing methodological innovations, adjusted GDP measures, novel aggregation techniques, subsequently adopted by official statistics. Non-governmental organisations produce 14% (9 indicators), typically emphasising dimensions neglected by official statistics. Commercial entities account for 5% (3 frameworks), bringing user-oriented design but facing sustainability questions.

Geographic scope follows a pronounced hierarchy. Global frameworks designed for international comparison represent 53% of the sample (35 indicators), typically covering 100 or more countries. These enable benchmarking across diverse contexts and identification of universal patterns, though standardisation requirements may obscure locally salient dimensions. National frameworks tailored to domestic circumstances comprise 38% (25 indicators), concentrated heavily in OECD and EU member states where statistical capacity, political interest in Beyond-GDP approaches, and resources for survey investment converge. Regional frameworks addressing EU territories or subnational areas account for 9% (6 indicators), enabling within-country comparison whilst respecting sub-national variation.

Methodological approaches reveal enduring tension between communicative simplicity and analytical richness. Composite indices aggregating multiple components into single scores remain dominant at 52% (34 indicators), facilitating communication through headline numbers comparable to GDP whilst enabling ranking that satisfies media interest and political accountability demands. Dashboard approaches presenting disaggregated indicators without aggregation represent 33% (22 frameworks), showing marked increase in recent national initiatives as developers recognise that aggregation obscures

the component-level information needed for targeted policy response. Survey-based measures relying primarily on subjective assessments comprise 11% (7 indicators), directly capturing lived experience but constrained by cost and frequency. Adjusted GDP measures modifying national accounts represent 5% (3 frameworks), maintaining connection to established economic measurement whilst attempting broader scope.

3.2 Domain Coverage: Concentration and Neglect

Systematic mapping of the 66 indicators against the 22-domain taxonomy reveals pronounced patterns of coverage concentration alongside systematic neglect of dimensions central to social service evaluation. Table 1 presents coverage rates across all domains, ordered by frequency of inclusion.

Table 1. Domain Coverage Across Beyond-GDP Indicators

Domain	Paper Sample (N=66)	Full Database (N=209)	Coverage Tier
Health	57 (86%)	162 (78%)	High
Knowledge & Skills	53 (80%)	150 (72%)	High
Material Wellbeing	51 (77%)	166 (79%)	High
Work & Job Quality	48 (73%)	133 (64%)	High
Economic Security	46 (70%)	131 (63%)	High
Environmental Sustainability	40 (61%)	105 (50%)	Moderate-High
Safety	36 (55%)	95 (45%)	Moderate
Housing	35 (53%)	90 (43%)	Moderate
Institutions	34 (52%)	99 (47%)	Moderate
Social Inclusion	31 (47%)	79 (38%)	Moderate
Trust & Social Connections	31 (47%)	71 (34%)	Moderate
Land & Ecosystem	29 (44%)	61 (29%)	Moderate
Subjective Wellbeing	29 (44%)	73 (35%)	Moderate
Climate	27 (41%)	73 (35%)	Moderate
Social Equity	25 (38%)	57 (27%)	Low-Moderate
Air Quality	24 (36%)	58 (28%)	Low-Moderate
Physical Capital	24 (36%)	54 (26%)	Low-Moderate
Energy Resources	21 (32%)	47 (22%)	Low-Moderate
Culture	18 (27%)	36 (17%)	Low
Water	17 (26%)	40 (19%)	Low
Leisure	17 (26%)	49 (23%)	Low
Care Services	11 (17%)	12 (6%)	Critical Gap

Note: Domains ordered by paper sample coverage. High ≥70%; Moderate-High 60-69%; Moderate 40-59%; Low-Moderate 30-39%; Low 20-29%; Critical Gap <20%. Full methodology and keyword dictionaries provided in Methodology Supplement.

Five domains achieve coverage of 70% or higher in the paper sample, forming a coherent human development cluster that reflects the enduring influence of frameworks established in the 1990s. Health appears in 57 of 66 indicators (86%), measured through diverse components spanning life expectancy, mortality rates across age groups and causes, healthy life years, healthcare access, service quality, and expenditure. This near-universal coverage reflects robust international measurement infrastructure through the WHO Global Health Observatory, established comparability of vital statistics across national systems, and widespread recognition that health outcomes represent fundamental constituents of human wellbeing regardless of theoretical orientation. Knowledge & Skills features in 53 indicators (80%), captured through educational attainment, literacy, standardised assessment performance, and participation in lifelong learning. Material Wellbeing appears in 51 indicators (77%), encompassing income, poverty, consumption, wealth, and economic sufficiency, dimensions readily measured through household surveys and national accounts. Work & Job Quality features in 48 indicators (73%), extending beyond simple employment rates to encompass job satisfaction, workplace conditions, work-life balance, and employment security, reflecting relatively recent recognition that employment quantity alone inadequately captures labour market wellbeing. Economic Security appears in 46 indicators (70%), addressing vulnerability and resilience through social protection coverage, financial buffers, and exposure to catastrophic expenditure. Together, these five domains demonstrate sophisticated measurement infrastructure developed over decades.

A second tier of domains demonstrates moderate coverage between 40% and 61%. Environmental Sustainability (61%) spans multiple sub-domains including air quality, climate change, ecosystem health, and resource use, though as discussed below these often appear in fragmented rather than integrated fashion. Safety (55%) captures crime rates, violence, and perceptions of personal security. Housing (53%) encompasses affordability, quality, adequacy, and tenure security. Institutions (52%) captures rule of law, government effectiveness, corruption control, and accountability mechanisms. Trust & Social Connections and Social Inclusion both appear in 47% of indicators. Subjective Wellbeing, capturing life satisfaction, happiness, and sense of meaning, features in 44%, concentrated in frameworks explicitly adopting experiential approaches post-Stiglitz-Sen-Fitoussi. Earlier generation indicators rarely included subjective measures, and acceptance remains incomplete with many statistical offices treating them as supplementary rather than core.

A third tier shows lower coverage in the 30-39% range. Social Equity (38%) addresses distribution through income inequality, equality of opportunity, and intergenerational mobility. Air Quality (36%), Physical Capital (36%), and Energy Resources (32%) round out this tier.

The lowest coverage domains reveal the systematic blind spots in Beyond-GDP measurement. Culture (27%) and Leisure (26%) receive minimal attention, reflecting productivist biases prioritising economic activity. Water (26%) shows surprisingly low coverage given SDG prominence, partly reflecting the geographic concentration of indicators in high-income contexts where universal water access reduces salience.

Most significantly for social service evaluation, Care Services appears in only 11 indicators (17%) in the paper sample. This figure alone would represent a striking gap given care services' essential role supporting children, persons with disabilities, and elderly individuals. But the most dramatic finding emerges from comparing the paper sample with the full database: Care Services coverage drops from 17% in the curated paper sample to just 6% (12 indicators) in the full database of 209 indicators, nearly a threefold decrease. This finding confirms that our paper sample, which includes frameworks meeting criteria for social content and service relevance, is considerably more favourable to care measurement than the broader Beyond-GDP field. The measurement gap reflects and reinforces broader patterns of care invisibility: when national progress frameworks fail to assess whether care needs are adequately met, care remains marginal to policy discourse regardless of rhetorical commitment to supporting vulnerable populations.

Service integration and coordination appears in fewer than seven indicators (approximately 11%), with most

frameworks measuring health, education, housing, and social protection as entirely separate domains without capturing whether services work together effectively for individuals with complex needs crossing sectoral boundaries. Prevention and early intervention proves virtually invisible at approximately 6% coverage. Frameworks readily capture crisis responses manifesting in hospitalisations, benefit claims, and emergency placements, whilst remaining largely unable to assess whether proactive strategies successfully prevent problems escalating.

3.3 Co-occurrence Patterns and Empirical Clustering

Moving beyond coverage frequencies, domain co-occurrence analysis examines which dimensions appear together across indicators, revealing the implicit conceptual structures underlying Beyond-GDP measurement. Table 2 presents co-occurrence rates for key domain pairs, organised by empirical cluster.

3.3 Co-occurrence Patterns and Empirical Clustering

Domain Pair	N	Rate	Cluster
Economic Foundations			
Material Wellbeing – Economic Security	46	70%	Economic
Material Wellbeing – Work Quality	43	65%	Economic
Economic Security – Work Quality	40	61%	Economic
Material Wellbeing – Housing	35	53%	Economic
Social Wellbeing			
Health – Knowledge & Skills	55	83%	Social
Health – Material Wellbeing	53	80%	Social/Economic
Health – Housing	34	52%	Social
Health – Social Inclusion	31	47%	Social
Environmental Sustainability			
Environmental Sust. – Air Quality	25	38%	Environmental
Environmental Sust. – Climate	22	33%	Environmental
Air Quality – Climate	20	30%	Environmental
Governance and Institutions			
Social Inclusion – Social Equity	27	41%	Governance
Social Inclusion – Institutions	25	38%	Governance
Institutions – Safety	23	35%	Governance
Undervalued Dimensions			
Care Services – Health	–	<20%	Isolated
Care Services – Social Inclusion	–	<20%	Isolated
Subjective Wellbeing – Health	18	27%	Isolated
Culture – Leisure	12	18%	Peripheral

Note: Values represent percentage of indicators containing both domains. Cluster assignments derived from hierarchical analysis of co-occurrence matrix.

The co-occurrence patterns validate a four-pillar structure frequently proposed in Beyond-GDP literature whilst revealing important variations in cluster coherence. Economic foundations demonstrates robust clustering: material wellbeing and economic security co-occur in 70% of indicators, material wellbeing and work quality in 65%, and economic security and work quality in 61%. Nearly half (44%) of all indicators include all three core economic domains together. Housing, whilst sometimes classified within social dimensions, shows stronger co-occurrence with economic domains (53% with material wellbeing) than with health or education, suggesting frameworks implicitly treat housing as economic circumstance rather than service provision. This tight clustering reflects theoretical coherence, income security, employment quality, and material living standards form an integrated whole where deficits in one dimension typically accompany deficits in others, and practical measurement infrastructure developed over decades for economic assessment.

Social wellbeing encompasses health, knowledge and skills, social connections, inclusion, and equity, though with somewhat lower internal coherence than the economic cluster. Health and knowledge co-occur in 83% of indicators, reflecting the Human Development Index template that established these as foundational wellbeing constituents. Health and housing co-occur in 52%, and health and inclusion in 47%. The social cluster demonstrates co-occurrence rates ranging from 47% to 83%, indicating recognised but incompletely integrated dimensions where measurement approaches continue evolving. The high overlap between health and education, and the moderate overlap of both with other social domains, suggests that social wellbeing functions as a meaningful pillar in practice whilst containing greater internal heterogeneity than the economic cluster.

Environmental sustainability shows moderate clustering at 30–38% co-occurrence among environmental sub-domains. Air quality and climate co-occur in 30% of indicators, environmental sustainability broadly defined with air quality in 38%. These lower rates compared with economic and social dimensions partially reflect environmental measurement's relative recency in Beyond-GDP frameworks and ongoing theoretical debate regarding whether environment constitutes wellbeing directly or merely enables it. The fragmentation also reflects measurement practicalities: climate change, air pollution, biodiversity, and resource use require different data sources and assessment methods, making comprehensive environmental coverage more demanding than aggregating income and employment statistics.

Governance and institutions demonstrates co-occurrence ranging from 33% to 41%, lower than economic foundations but comparable to social wellbeing. Social inclusion and equity co-occur in 41% of indicators, institutions and safety in 35%. This cluster addresses power relations, procedural fairness, and enabling conditions for agency, dimensions increasingly recognised as relevant to wellbeing but integrated less consistently than economic or health outcomes.

The most significant finding for social service evaluation concerns domains demonstrating weak integration with these established clusters. Care Services appears alongside Health – the domain with which it shares most obvious conceptual affinity, in fewer than 20% of indicators. Care and Social Inclusion similarly co-occur in fewer than 20% of cases. These rates stand in stark contrast to the 50–83% co-occurrence characterising core economic and social domains. Subjective Wellbeing shows 27% co-occurrence with Health, lower than might be expected given both dimensions' focus on individual experience. Culture and Leisure co-occur with each other at only 18%, and both show similarly weak integration with established pillars.

This pattern of weak integration confirms that Care Services, Subjective Wellbeing, Culture, and Leisure function as orphaned dimensions in Beyond-GDP measurement, included sporadically in individual frameworks but not incorporated into the conceptual architecture that structures the field. The implication is profound: addressing these gaps requires deliberate framework design rather than expectation that expanded measurement will organically incorporate currently neglected dimensions. Frameworks treating care as an afterthought or optional addition will

perpetuate current patterns; only those explicitly architecting care, experience quality, and temporal dimensions into core structure will achieve comprehensive coverage.

3.4 Approaches to Social Service Measurement

Classification of indicators by their underlying measurement logic reveals four archetypal approaches to conceptualising social services, each with characteristic strengths and limitations for evaluation purposes. These types are not mutually exclusive, many frameworks combine elements, but distinguishing the underlying logics clarifies what different measurement approaches can and cannot capture.

System performance frameworks, representing approximately 26% of the sample, treat social services as systems to be optimised according to efficiency and effectiveness criteria. Examples include the EU Social Scoreboard, OECD social protection statistics, and various health system performance indices. These frameworks typically measure coverage rates indicating what proportion of eligible populations receive services, access metrics such as waiting times and geographic proximity, efficiency ratios relating inputs to outputs, and processing indicators capturing speed and accuracy of service administration. The system performance approach offers clear accountability metrics suitable for managerial oversight and enables identification of bottlenecks constraining service delivery. Its limitations centre on the reduction of complex services to relatively simple metrics potentially amenable to gaming, difficulty capturing quality dimensions that matter to users but resist quantification, and systematic focus on service characteristics rather than outcomes in users' lives.

Human development frameworks constitute the largest group at approximately 35%, treating services as investments in human capabilities following the theoretical tradition established by Amartya Sen and operationalised through UNDP measurement. The Human Development Index and its variants, the Social Progress Index, and most national wellbeing dashboards fall within this category. These frameworks assess services indirectly through population outcomes, life expectancy, educational attainment, literacy rates, that services influence alongside many other factors. The approach focuses appropriately on ultimate goals rather than intermediate processes and enables international comparison through standardised outcome metrics. However, human development measurement struggles with attribution: observed capability levels reflect not only service provision but family circumstances, economic conditions, peer influences, and individual choices. Long lag times between service delivery and capability formation further complicate accountability when outcomes manifest years or decades after intervention.

Subjective wellbeing frameworks, comprising approximately 15% of the sample, prioritise user experience and satisfaction as the ultimate criterion for assessing whether services generate value. The World Happiness Report, Australian Unity Wellbeing Index, and Better Life Index life satisfaction components exemplify this approach. Subjective frameworks capture what system performance and capability approaches may miss: whether service provision translates into experienced improvement in quality of life, and whether delivery processes respect dignity and autonomy in ways that matter to users. Limitations include adaptation effects whereby expectations adjust to service levels, cultural variation in response styles complicating comparison, and potential difficulty translating satisfaction scores into actionable guidance regarding which specific service dimensions require improvement.

Rights and entitlements frameworks, representing approximately 24%, assess services against normative standards derived from legal obligations or ethical principles. The Sustainable Development Goals, human rights monitoring frameworks, and various equality indices evaluate whether legal entitlements are fulfilled, whether access is non-discriminatory across protected characteristics, whether accountability mechanisms enable recourse when entitlements are denied, and whether minimum acceptable standards are achieved. The rights approach provides unambiguous assessment criteria independent of political preferences, and legal backing strengthens claims for service adequacy. Limitations include binary pass-fail logic potentially insensitive to gradations and incremental improvement, difficulty

establishing thresholds simultaneously ambitious and realistic across contexts with vastly different resources, and tension between the rhetoric of immediate obligation and pragmatic acknowledgment of resource constraints.

None of these approaches proves comprehensively adequate for social service evaluation. System performance captures efficiency but not ultimate value. Human development assesses outcomes but struggles with attribution. Subjective wellbeing centres user experience but may lack specificity regarding service dimensions requiring attention. Rights frameworks provide normative clarity but resist empirical gradation. Comprehensive evaluation requires integration across approaches, combining efficiency assessment with outcome measurement, user experience with normative standards, within frameworks capable of maintaining conceptual coherence whilst spanning these distinct evaluative logics.

Table 3. Methodological Characteristics of Beyond-GDP Indicators

Characteristic	Distribution
Data Sources	
Administrative data	59 (89%)
Survey data	41 (62%)
Mixed sources	31 (47%)
Update Frequency	
Annual	42 (64%)
Multi-year or irregular	18 (27%)
Quarterly or more frequent	2 (3%)
Discontinued or unclear	4 (6%)
Publication Lag	
Median lag from reference period	18 months
Comparison: GDP quarterly accounts	6–8 weeks
Weighting Approaches (composites, n=34)	
Equal weighting	18 (53%)
Statistical weighting	6 (18%)
Expert weighting	8 (24%)
Implicit (geometric mean)	2 (6%)
Equity Disaggregation	
Gender	37 (56%)
Age	31 (47%)
Geographic	45 (68%)
Income/wealth	34 (52%)
Disability	8 (12%)
Migration status	6 (9%)
Ethnicity/race	12 (18%)
Intersectional analysis	8 (12%)

3.5 Methodological Characteristics

The utility of these frameworks does not solely depend on the conceptual orientation. How indicators are constructed, updated and disaggregated shapes whether they can inform the responsive and equity-focused assessment social services require. The methodological aspects of Beyond-GDP frameworks reveal opportunities and barriers for such application. Table 3 summarises key characteristics across the 66 indicators.

Data source patterns reveal characteristic strengths and gaps. Administrative data appears in 89% of indicators, leveraging statistics routinely collected through service delivery, benefit administration, and population registration systems. Administrative sources provide comprehensive population coverage, regular availability enabling frequent updates, and low marginal cost once collection infrastructure exists. However, administrative data captures transactions and legal categories rather than experiences and outcomes, service contacts and benefit payments rather than whether needs were met or lives improved. Survey data features in 62% of frameworks, enabling direct assessment of subjective experience, quality perceptions, and dimensions absent from administrative records. Survey approaches face cost constraints limiting frequency and sample size, particularly for granular geographic or demographic disaggregation. Mixed-source frameworks combining administrative and survey data appear in 47% of indicators, offering optimal potential but demanding complex coordination across collection systems operated by different organisations.

Update frequency and timeliness critically affect policy utility. Annual updates characterise 64% of frameworks, enabling regular monitoring and alignment with budget cycles. Multi-year or irregular updates apply to 27%, often reflecting dependence on infrequent census data or resource-intensive primary collection. Only 3% achieve quarterly or more frequent updates, both exploiting monthly administrative series enabling near real-time reporting. Median publication lag from reference period to data availability spans 18 months across frameworks reporting this information. This contrasts starkly with GDP's quarterly accounts released within six to eight weeks of period end. The timeliness gap partially explains GDP's persistent dominance in policy discourse despite its recognised limitations: when policy cycles demand current information, Beyond-GDP frameworks providing 18-month-old data face structural disadvantage regardless of their superior conceptual coverage.

Aggregation approaches among composite indices reveal fundamental philosophical choices with implications for how different dimensions are weighted. Equal weighting appears in 53% of composites, avoiding contentious value judgments but implicitly assuming all domains contribute equally to progress regardless of their actual importance or the severity of deficits. Statistical weighting through methods such as principal components analysis features in 18%, allowing weights to emerge from observed variation but potentially prioritising measurability over importance. Expert weighting appears in 24%, incorporating professional judgment but risking elite bias when expert priorities diverge from those of affected populations. The growing preference for dashboard approaches avoiding aggregation entirely (33% of total sample) reflects recognition that weighting choices prove contentious and that preserving component-level information enables more targeted policy response than aggregate scores obscuring dimensional performance. Beyond construction choices, the capacity of frameworks to illuminate distributional patterns proves equally consequential for social service evaluation.

Equity and disaggregation capabilities determine whether frameworks can assess distributional concerns central to social service evaluation. Gender disaggregation appears in 56% of indicators, age in 47%, and geographic breakdown in 68%, relatively high rates reflecting established statistical practice. Income or wealth disaggregation features in 52%, enabling assessment of whether progress reaches disadvantaged populations. However, disaggregation by characteristics directly relevant to service need remains far less common. Disability disaggregation appears in only 12% of frameworks, migration status in 9%, and ethnicity or race in 18%. The near-absence

of disability disaggregation proves particularly striking given that persons with disabilities represent a population with intensive service needs and heightened vulnerability to access barriers and quality deficits. Most critically, intersectional analysis combining multiple characteristics to examine compounding disadvantages, elderly women with disabilities, migrants with mental health difficulties, appears in just 12% of frameworks. This severe limitation prevents examination of how disadvantages compound in ways that siloed single-characteristic analysis cannot capture.

Taken together, these methodological characteristics reveal that current measurements lack the capacity to assess differentiated needs. The reliance on administrative data privileges coverage over experience quality, publication lags constraint policy responsiveness and limited intersectional disaggregation obscure the compounding vulnerabilities that social services are specifically designed to address. These are not merely technical limitations but structural constraints shaping what questions Beyond-GDP frameworks can and cannot answer regarding social service adequacy.

4. Discussion

The findings reveal a paradox at the heart of Beyond-GDP measurement. Five decades of sustained effort have produced sophisticated assessment of health outcomes, educational attainment, material living standards, and increasingly environmental sustainability, dimensions that GDP systematically neglects. Yet the frameworks designed to capture what GDP misses themselves contain systematic blind spots regarding social services, care provision, and the quality of service experience. This section interprets these findings and considers their implications for developing comprehensive social service evaluation frameworks.

4.1 Understanding the Measurement Gaps

The systematic underrepresentation of care services, service integration, and prevention reflects structural biases in measurement practice rather than technical impossibility or conceptual ambiguity regarding their importance. Care work remains invisible across multiple measurement systems – GDP counts market childcare whilst ignoring the far larger volume of unpaid family care; labour statistics undercount informal care arrangements; and now Beyond-GDP frameworks prove to perpetuate rather than correct this invisibility. The devaluation of care in progress measurement reflects its broader devaluation in economic and social accounting: work performed predominantly by women, disproportionately for the benefit of vulnerable populations, within relationships rather than market transactions.

The comparison between paper sample and full database coverage proves particularly revealing. Care Services drops from 17% coverage in our curated sample of 66 indicators to just 6% in the full database of 209 indicators, nearly a threefold decrease. This pattern confirms that frameworks meeting our criteria for social content and service relevance are substantially more likely to include care measurement than the broader field. The implication is sobering: the broader Beyond-GDP movement has largely failed to incorporate care dimensions, and only frameworks with explicit social service orientation approach even minimal care coverage.

Service integration presents measurement challenges of a different nature. Statistical systems have developed along administrative lines, with health ministries collecting health statistics, education departments gathering educational data, and social security administrations tracking benefit receipt. This siloed architecture proves adequate for assessing sectoral performance but structurally incapable of capturing whether services coordinate effectively for individuals whose needs cross departmental boundaries. No established consensus exists regarding how to operationalise integration as a measurable construct distinct from the performance of component services, and no obvious data source captures coordination quality across organisational interfaces.

Prevention measurement faces perhaps the most fundamental conceptual challenge. Successful prevention produces non-events, hospitalisations that did not occur, crises that never escalated, benefit claims that were never filed. Measuring prevention requires counterfactual reasoning regarding what would have happened absent intervention, attribution frameworks distinguishing preventive effects from other influences, and willingness to value problems averted rather than only crises managed. These requirements conflict with measurement systems designed to count events rather than estimate their absence.

The co-occurrence analysis provides important additional insight into the persistence of these gaps. Care Services, Subjective Wellbeing, Culture, and Leisure do not simply appear infrequently, they appear without consistent relationship to other measured domains. The below-20% co-occurrence between care and other social dimensions compares with above 70% among established pillar domains. This isolation indicates that care has not been integrated into the conceptual architecture structuring Beyond-GDP measurement; it remains an optional addition that individual framework developers may or may not include rather than a recognised constituent that comprehensive frameworks should address. The implication is that expanded measurement will not automatically incorporate care dimensions. Deliberate architectural decisions are required to ensure care, experience quality, and temporal dimensions achieve integration comparable to economic and health outcomes.

4.2 Implications for Framework Development

These findings carry significant implications for developing comprehensive social service evaluation frameworks. Most fundamentally, existing Beyond-GDP frameworks cannot serve as ready-made templates without substantial modification. Despite their progress beyond GDP's limitations, they do not adequately address care provision, service coordination, or prevention, dimensions that should be central to any comprehensive assessment of social service systems. Framework development must deliberately incorporate these undervalued dimensions rather than assuming they will emerge through incremental expansion of existing approaches.

Methodological pluralism appears necessary rather than optional. The four archetypal approaches identified – system performance, human development, subjective wellbeing, and rights assessment – each capture dimensions that others miss. System performance frameworks provide the efficiency and access metrics needed for service management but cannot assess whether services ultimately improve lives. Human development frameworks focus appropriately on outcomes but struggle with attribution when services represent one influence among many on health, education, and economic circumstances. Subjective wellbeing approaches centre user experience but may lack specificity regarding which service dimensions require improvement. Rights frameworks provide normative clarity but resist the empirical gradation needed to track incremental progress. Comprehensive evaluation requires integration across these traditions, combining their respective strengths whilst compensating for individual limitations.

Disaggregation capabilities require substantial enhancement. The finding that only 12% of frameworks support intersectional analysis, and that disability disaggregation appears in just 12% of indicators, represents a critical gap for social service evaluation. Services explicitly target populations facing multiple disadvantages; elderly persons with disabilities, migrants experiencing mental health difficulties, families facing compound deprivation across employment, housing, and child welfare. Evaluation frameworks incapable of examining compounding vulnerabilities cannot adequately assess whether services reach those with greatest need or whether interventions address intersecting rather than isolated disadvantages.

The timeliness gap between Beyond-GDP frameworks and GDP creates structural disadvantage that framework developers must address. When policy decisions require current information and GDP provides quarterly accounts within weeks whilst alternative measures offer annual data released after

18 months, GDP will continue dominating policy discourse regardless of its conceptual limitations. Investment in nowcasting techniques, high-frequency administrative indicators, and preliminary estimates accepting reduced precision for improved timeliness could significantly enhance policy relevance.

Finally, the concentration of sophisticated national frameworks within OECD contexts raises questions regarding global applicability. The frameworks developed in settings with universal public services, comprehensive administrative data infrastructure, and substantial survey investment may not adequately address the challenges of social service evaluation in contexts where provision models differ substantially, informal arrangements predominate, and statistical capacity is more constrained. Developing frameworks applicable across diverse welfare state configurations requires deliberate attention to adaptability rather than assumption that approaches proven in high-income settings will transfer without modification.

5. Conclusion

This meta-analysis of 66 Beyond-GDP indicators developed between 1972 and 2023 provides the first systematic assessment of how alternative progress measures address social services. The findings reveal both remarkable achievement and persistent limitation. Health (86%), Knowledge & Skills (80%), Material Wellbeing (77%), Work & Job Quality (73%), and Economic Security (70%) form a coherent human development cluster demonstrating sophisticated measurement reflecting decades of methodological development and institutional investment. Domain co-occurrence analysis validates an empirically grounded four-pillar structure (economic foundations, social wellbeing, environmental sustainability, and governance) with robust clustering among core dimensions.

Still domains central to social service provision remain systematically neglected. Care Services appears in only 17% of indicators in our sample, and just 6% in the full database of 209 indicators, despite its essential role supporting vulnerable populations. This nearly threefold decrease from curated sample to full database represents the most dramatic coverage gap in the entire analysis, confirming that the broader Beyond-GDP field has largely failed to incorporate care measurement. Service integration features in approximately 11% and prevention in approximately 6%. These gaps prove structural rather than incidental: co-occurrence analysis reveals care achieving below 20% integration with established measurement clusters compared with above 70% among core pillar domains. Methodological limitations compound these coverage gaps. Only 12% of frameworks support intersectional analysis, disability disaggregation appears in just 12% of indicators, and publication lags averaging 18 months limit utility for responsive policymaking.

The Beyond-GDP movement has demonstrated that comprehensive progress measurement is both conceptually sound and practically feasible. The proliferation of national wellbeing dashboards across OECD member states, the global architecture of the Sustainable Development Goals, and the increasing integration of wellbeing evidence with policy processes represent genuine advances beyond GDP's limitations. Yet systematic blind spots persist regarding social services constituting major welfare state investments. OECD countries dedicate over one-fifth of GDP to social expenditure spanning health, pensions, family support, and disability services (OECD, 2024), yet measurement of service quality, coordination effectiveness, and prevention impact remains underdeveloped.

Addressing these gaps requires deliberate framework design rather than expectation that expanded measurement will organically incorporate currently neglected dimensions. The weak integration of care, subjective wellbeing, and temporal dimensions with established measurement pillars indicates that incremental extension of existing approaches will

perpetuate rather than correct current limitations. Comprehensive social service evaluation demands frameworks explicitly architecting care provision, service coordination, prevention effectiveness, and user experience into core structure alongside economic circumstances and population outcomes.

These findings establish an evidence base for developing such frameworks, though several limitations must be acknowledged. The restriction to English-language documentation may have excluded innovative non-Anglophone approaches, particularly from the Global South, and approximately 30% of indicators lacked complete component inventories. Subsequent work will extend the analysis through dominance assessment identifying which existing indicators best address social service dimensions, providing guidance for practitioners and policymakers seeking to strengthen evidence-informed service evaluation. When measurement systems reflect what societies genuinely value, including the care that enables dignity and independence for vulnerable populations, they enable more informed choices regarding resource allocation and institutional design affecting human flourishing.

Acknowledgments

This research was developed within the framework of the Horizon Europe project: Building Economic, Needs-Based and Environmental evaluation Frameworks for Inclusive Transformation of Social services (BENEFITS), Grant Agreement No. 101136407. The authors thank Nikos Avgeris, Eleni Vossou, George Strofyllas, and Nilufer Sari Aslam for research assistance, and Pinar Cakiroglu for review comments on earlier drafts.

References

Beck, U. (1999). World Risk Society. Polity Press.

Cobb, C., Halstead, T., & Rowe, J. (1995). The Genuine Progress Indicator: Summary of data and methodology. Redefining Progress.

Costanza, R., Kubiszewski, I., Giovannini, E., Lovins, H., McGlade, J., Pickett, K. E., Ragnarsdóttir, K. V., Roberts, D., De Vogli, R., & Wilkinson, R. (2014). Development: Time to leave GDP behind. Nature, 505(7483), 283–285.

Daly, H. E., & Cobb, J. B. (1989). For the Common Good: Redirecting the Economy toward Community, the Environment, and a Sustainable Future. Beacon Press.

Fleurbaey, M., & Blanchet, D. (2013). Beyond GDP: Measuring Welfare and Assessing Sustainability. Oxford University Press.

Hoekstra, R. (2019). Replacing GDP by 2030: Towards a Common Language for the Well-being and Sustainability Community. Cambridge University Press.

ILO. (2018). Care Work and Care Jobs for the Future of Decent Work. International Labour Organization.

ILO. (2022). Global Care Policy Index. International Labour Organization.

ILO. (2024). World Employment and Social Outlook: The Care Economy. International Labour Organization.

Johnston, D. (1985). Measuring social welfare. Economic Council of Canada.

Klasen, S. (2007). Gender-related indicators of well-being. In M. McGillivray (Ed.), Human Well-being: Concept and Measurement (pp. 167–192). Palgrave Macmillan.

Kubiszewski, I., Costanza, R., Anderson, S., & Sutton, P. (2025). A comprehensive database of wellbeing indicators: Classification and analysis. Ecological Economics [forthcoming].

Kuznets, S. (1934). National Income, 1929–1932. National Bureau of Economic Research.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine, 6(7), e1000097.

Nordhaus, W. D., & Tobin, J. (1972). Is growth obsolete? In Economic Research: Retrospect and Prospect, Volume 5: Economic Growth (pp. 1–80). National Bureau of Economic Research.

OECD. (2013). OECD Guidelines on Measuring Subjective Well-being. OECD Publishing.

OECD. (2024). Society at a Glance 2024: OECD Social Indicators. OECD Publishing. https://doi.org/10.1787/918d8db3-en

Rum, A., et al. (2024). Domain taxonomy for wellbeing indicators: A semantic clustering approach. MERGE Project Working Paper.

Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress. Paris.

Stockhammer, E., Hochreiter, H., Obermayr, B., & Steiner, K. (1997). The Index of Sustainable Economic Welfare (ISEW) as an alternative to GDP in measuring economic welfare. Ecological Economics, 21(1), 19–34.

UN Women. (2016). Redistributing Unpaid Care and Sustaining Quality Care Services: A Prerequisite for Gender Equality. Policy Brief No. 5.

Appendix A: Selection Process Summary

Identification: Initial universe of 213 wellbeing indicators from Kubiszewski et al. (2025) database supplemented by independent database searches and snowball sampling from major reviews.

Screening Stage 1 – Binary Criteria: Criterion 1 (Temporal relevance: active 2000–2024 or foundational pre-2000) excluded 26 indicators, yielding 187. Criterion 2 (Scale: ≥3 countries, EU/OECD coverage, or international reference status) excluded 63, yielding 124. Criterion 3 (Public English documentation) excluded 6, yielding 118. Criterion 4 (Multi-dimensional: ≥3 domains) excluded 29, yielding 89 indicators meeting all binary criteria.

Screening Stage 2 – Relevance Scoring: Criterion 5 (Social content prominence, 0–3 scale) and Criterion 6 (Service domain relevance, 0–3 scale) applied to 89 indicators. Combined score threshold of ≥3 excluded 17 indicators, yielding 72.

Recency Verification: Exclusion of frameworks discontinued >7 years before 2024 removed 6 indicators (Environmental Sustainability Index, Green GDP China, National Accounts of Wellbeing, Where-to-be-born Index, Measures of Australia's Progress, Sustainable Society Index). Human Poverty Index retained as predecessor to active MPI.

Final Sample: 66 Beyond-GDP indicators spanning 1972–2023 included for analysis.

Appendix B: Complete Indicator Inventory

Table B1 presents the 66 Beyond-GDP indicators included in this meta-analysis, ordered by year of introduction.

Table B1. Included Indicators

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
1	Genuine Progress Indicator (GPI)	Academic (Redefining Progress)	Adjusted GDP	1995	National (USA, adapted globally)	27	Irregular
2	Human Development Index (HDI)	UN Development Programme	Composite	1990	Global (191 countries)	4	Annual
3	Inequality-adjusted HDI (IHDI)	UN Development Programme	Composite	2010	Global (151 countries)	5	Annual
4	OECD Better Life Index (BLI)	OECD	Dashboard	2011	OECD members (38 countries)	24	Biennial
5	Inclusive Wealth Index (IWI)	UN Environment Programme	Composite	2012	Global (140+ countries)	17	Irregular (4-year)
6	Happy Planet Index	New Economics Foundation (NGO)	Composite	2006	Global (140+ countries)	3	Irregular
7	Index of Sustainable Economic Welfare (ISEW)	Academic (Daly & Cobb)	Adjusted GDP	1989	National (multiple adaptations)	8	Irregular
8	Legatum Prosperity Index	Legatum Institute (NGO)	Composite	2007	Global (167 countries)	67	Annual
9	Canadian Index of Wellbeing (CIW)	University of Waterloo	Dashboard	2011	National (Canada)	62	Biennial
10	City Prosperity Index (CPI)	UN-Habitat	Composite	2012	Global cities (400+)	62	Irregular

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
11	Sustainable Wellbeing Index (SWI)	Academic (multiple)	Composite	2010	National (various)	3	Irregular
12	Genuine Savings/ Adjusted Net Savings (ANS)	World Bank	Adjusted GDP	1999	Global (140+ countries)	7	Annual
13	Gallup-Sharecare Wellbeing Index	Gallup (Business)	Survey	2008	National (USA)	42	Continuous
14	AARP Liveability Index	AARP (NGO)	Composite	2015	National (USA communities)	59	Annual
15	National Wellbeing: Life Satisfaction	UK Office for National Statistics	Dashboard	2010	National (UK)	6	Quarterly
16	Regional Quality of Life Index	Eurostat	Composite	2015	EU NUTS-2 regions (263)	17	Biennial
17	Australian Unity Wellbeing Index	Australian Centre on Quality of Life	Survey	2001	National (Australia)	16	Biannual
18	Social Progress Index	Social Progress Imperative (NGO)	Composite	2013	Global (169 countries)	60	Annual
19	Multidimensional Poverty Index (MPI)	UNDP & Oxford Poverty & Human Development Initiative	Composite	2010	Global (111 countries)	10	Annual
20	Global Competitiveness Index	World Economic Forum	Composite	2004	Global (141 countries)	12	Annual
21	Gender Development Index (GDI)	UN Development Programme	Composite	1995	Global (167 countries)	5	Annual
22	Positive Peace Index (PPI)	Institute for Economics & Peace	Composite	2009	Global (163 countries)	24	Biennial
23	European Quality of Life Survey	Eurofound	Survey	2003	EU28 + candidate countries	12	4-year cycle
24	Quality of Life Index	Numbeo (Business)	Composite	2009	Global cities (continuous)	26	Continuous
25	Bloomberg Healthiest Countries Index	Bloomberg (Business)	Composite	2019	Global (169 countries)	10	Irregular
26	Living Standards Framework (LSF)	New Zealand Treasury	Dashboard	2011	National (New Zealand)	21	Annual
27	Thriving Places Index (TPI)	Local Trust (UK NGO)	Composite	2017	National (England localities)	22	Annual
28	The Good Country Index	Good Country (NGO)	Composite	2014	Global (169 countries)	14	Irregular
29	European Social Survey (ESS)	European Research Infrastructure	Survey	2002	European (30+ countries)	19	Biennial
30	Mercer Quality of Living	Mercer (Business)	Composite	1999	Global cities (230+)	39	Annual

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
31	Global Liveability Index	Economist Intelligence Unit	Composite	2002	Global cities (173)	33	Annual
32	Human Sustainable Development Index	Academic (Togtokh & Gaffney)	Composite	2010	Global (151 countries)	9	Irregular
33	Gender Inequality Index (GII)	UN Development Programme	Composite	2010	Global (162 countries)	10	Annual
34	World Values Survey (WVS)	World Values Survey Association	Survey	1981	Global (120+ countries)	295	5-year waves
35	European Values Study (EVS)	EVS Foundation	Survey	1981	European (47+ countries)	70	9-year waves
36	Human Poverty Index (HPI)	UN Development Programme	Composite	1997	Global (replaced by MPI 2010)	5	Annual
37	Gross National Happiness (GNH) Index	Government of Bhutan	Composite	1972	National (Bhutan)	33	5-year
38	Inclusive Development Index	World Economic Forum	Composite	2017	Global (103 countries)	9	Annual
39	Sustainable Development Goals Index (SDGI)	Sustainable Development Solutions Network	Composite	2016	Global (193 countries)	97	Annual
40	UK Measures of National Wellbeing	UK Office for National Statistics	Dashboard	2010	National (UK)	46	Annual
41	Measuring What Matters Framework	Statistics Canada	Dashboard	2021	National (Canada)	81	Annual
42	Measuring Ireland's Progress	Central Statistics Office Ireland	Dashboard	2003	National (Ireland)	43	Annual
43	Finland's Findicators	Statistics Finland	Dashboard	2009	National (Finland)	83	Annual
44	Israel Wellbeing Indicators	Israel Central Bureau of Statistics	Dashboard	2015	National (Israel)	19	Annual
45	Quality of Life Indicators in Slovenia	Statistical Office Slovenia	Dashboard	2015	National (Slovenia)	16	Annual
46	Quality of Life Indicators in Korea	Statistics Korea	Dashboard	2011	National (South Korea)	8	Annual
47	Italy BES (Equitable & Sustainable Wellbeing)	Istat (Italian National Statistical Institute)	Dashboard	2010	National (Italy)	12	Annual
48	Belgium Sustainable Development Indicators	Federal Planning Bureau Belgium	Dashboard	2000	National (Belgium)	7	Annual
49	Chile Social Wellbeing Survey	Ministry of Social Development Chile	Survey	2015	National (Chile)	48	Biennial
50	Canada's Quality of Life Framework	Statistics Canada	Dashboard	2021	National (Canada)	8	Annual

#	Indicator Name	Developer Institution	Type	Year	Geographic Scope	Compon. (N)	Update Freq
51	European Social Progress Index	European Commission	Composite	2016	EU regions (272 NUTS-2)	11	Biennial
52	Quality of Life in Norway	Statistics Norway	Dashboard	2006	National (Norway)	8	Annual
53	Quality of Life Indicators in Spain	Spanish National Statistics Institute	Dashboard	2013	National (Spain)	37	Annual
54	Iceland Wellbeing Indicators	Statistics Iceland	Dashboard	2019	National (Iceland)	17	Annual
55	Indicators Aotearoa New Zealand	Stats NZ & Treasury	Dashboard	2018	National (New Zealand)	43	Annual
56	Switzerland MONET 2030	Swiss Federal Statistical Office	Dashboard	2003	National (Switzerland)	98	Annual
57	Netherlands Monitor of Wellbeing	Statistics Netherlands	Dashboard	2013	National (Netherlands)	21	Annual
58	Wellbeing in Germany	Federal Statistical Office Germany	Dashboard	2016	National (Germany)	14	Biennial
59	Doughnut Economics	Academic (Kate Raworth)	Dashboard	2017	City-level applications	Variable	Variable
60	Comprehensive Wealth (World Bank)	World Bank	Composite	2021	Global (146 countries)	17	Irregular
61	Ecological Footprint	Global Footprint Network	Composite	1992	Global (190+ countries)	6	Annual
62	Environmental Performance Index (EPI)	Yale & Columbia Universities	Composite	2006	Global (180 countries)	40	Biennial
63	Country Policy & Institutional Assessment (CPIA)	World Bank	Composite	1977	Low-income countries (78)	50	Annual
64	Ibrahim Index of African Governance (IIAG)	Mo Ibrahim Foundation	Composite	2007	African countries (54)	42	Annual
65	Global Gender Gap Index	World Economic Forum	Composite	2006	Global (146 countries)	16	Annual
66	World Happiness Report	UN Sustainable Development Solutions Network	Composite	2012	Global (143 countries)	10	Annual

Appendix C: Domain Taxonomy

Table C1 presents the 22-domain taxonomy used for systematic coverage mapping, adapted from the MERGE project classification (Rum et al., 2024) with modifications for social service evaluation relevance.

Domain	Definition and Scope
1. Health	Physical and mental health outcomes, healthcare access, morbidity, mortality, healthy life expectancy
2. Knowledge and Skills	Educational attainment, literacy, numeracy, cognitive skills, lifelong learning participation
3. Material Wellbeing	Income, consumption, wealth, poverty, material living standards and sufficiency
4. Work and Job Quality	Employment status, job security, working conditions, satisfaction, work-life balance
5. Economic Security	Financial resilience, social protection coverage, vulnerability to shocks, benefit adequacy
6. Housing	Affordability, quality, adequacy, tenure security, homelessness, housing conditions
7. Social Inclusion	Participation in social, economic, cultural, and political life; integration; belonging
8. Social Equity	Distribution of resources and opportunities; inequality; intergenerational mobility
9. Environmental Sustainability	Aggregate environmental quality; natural capital; ecological limits; sustainability
10. Institutions and Governance	Rule of law, government effectiveness, corruption control, accountability, voice
11. Safety and Security	Personal security, crime rates, violence, perceived safety, freedom from harm
12. Trust and Social Connections	Interpersonal trust, institutional trust, social networks, relationships, loneliness
13. Subjective Wellbeing	Life satisfaction, happiness, affect balance, sense of meaning and purpose
14. Care Services	Childcare, eldercare, disability support provision; informal care; carer wellbeing
15. Air Quality	Ambient air pollution (PM2.5, PM10, NO2, O3), exposure, indoor air quality
16. Water	Access to safe drinking water, sanitation, water quality, water stress
17. Climate	Greenhouse gas emissions, climate change impacts, mitigation, adaptation
18. Land and Ecosystems	Land use, biodiversity, ecosystem health, protected areas, natural habitats
19. Energy Resources	Energy access, efficiency, renewable share, fuel poverty, energy security
20. Physical Capital	Infrastructure quality, transport, telecommunications, public facilities
21. Culture and Heritage	Cultural participation, arts access, heritage preservation, cultural vitality
22. Leisure and Time Use	Time allocation, leisure quality, work hours, time poverty, work-life balance

Institute for Global Prosperity

Research at the UCL Institute for Global Prosperity aims to generate new insights about sustainable and inclusive prosperity and provide new models for developing and interpreting evidence.

Underlying our research is a rethinking of what we mean by prosperity. Prosperity must mean enabling people to flourish in ways beyond financial growth – and doing so equitably and sustainably, for humankind and the planet. We work with businesses, non-governmental organisations and citizens to produce interdisciplinary methodologies and problem-focused research.

For more information about our wide range of current projects and our innovative Masters and PhD programmes please see: https://www.ucl.ac.uk/bartlett/igp/.

Contact

igp@ucl.ac.uk
@glo_pro

www.prosperity-global.org
www.ucl.ac.uk/bartlett/igp/

UCL