European Healthy Cities evaluation: conceptual framework and methodology

Evelyne de Leeuw1,*, Geoff Green2, Mariana Dyakova3, Lucy Spanswick1, and Nicola Palmer1

1La Trobe University Melbourne, Melbourne, Australia, 2Sheffield Hallam University, Sheffield, UK, and 3Warwick University, Coventry, UK

*Corresponding author. E-mail: e.deleeuw@latrobe.edu.au

Summary

This paper presents the methodology, programme logic and conceptual framework that drove the evaluation of the Fifth Phase of the WHO European Healthy Cities Network. Towards the end of the phase, 99 cities were designated progressively through the life of the phase (2009–14). The paper establishes the values, systems and aspirations that these cities sign up for, as foundations for the selection of methodology. We assert that a realist synthesis methodology, driven by a wide range of qualitative and quantitative methods, is the most appropriate perspective to address the wide geopolitical, demographic, population and health diversities of these cities. The paper outlines the rationale for a structured multiple case study approach, the deployment of a comprehensive questionnaire, data mining through existing databases including Eurostat and analysis of management information generation tools used throughout the period. Response rates were considered extremely high for this type of research. Non-response analyses are described, which show that data are representative for cities across the spectrum of diversity. This paper provides a foundation for further analysis on specific areas of interest presented in this supplement.

Key words: healthy cities, model, evaluation of healthy cities network, evaluation methodology

INTRODUCTION

The global Healthy Cities movement was launched in 1986 and has thrived since. The European Healthy Cities and Networks, led by the World Health Organisation Regional Office for Europe (WHO EURO), have adopted a vision and a set of core attributes to work towards improving urban health and environment. Each 5 years (a Phase), the network has focused on different priority themes and engaged into relevant projects to achieve these.

Evaluation efforts until now have yielded significant evidence that being a ‘Healthy City’—belonging to the network, working within its values and approaches and putting health as a priority on the local agenda—constitutes agency for change, and designated Healthy Cities as well as (national and language) networks of Healthy Cities drive urban health innovation across Europe.

PROGRAMME LOGIC AND METHODOLOGICAL PRELIMINARIES

The evaluation of Phase V (2009–13) of the WHO European Healthy Cities Network was initiated in 2012 drawing on (i) past evaluation efforts and deliverables and (ii) current insights into the study of complex or
‘wicked’ health policy issues (De Leeuw, 2009; Ferlie et al., 2011).

Earlier phases of the WHO European Healthy Cities Network have been evaluated as well, and findings have been published. The first phase was explored by Draper et al. (Draper et al., 1993) harvesting commonalities among the designated cities in developing infrastructures and values. The evaluation approach was largely qualitative and resembled a ‘grounded theory’ approach (Glaser and Strauss, 2009). For Phase II, a research collaboration was established between the European Union and research institutions in Britain, The Netherlands and Italy (Berkeley and Humphreys, 1998; De Leeuw et al, 1998; Capello, 2000) studying a selection of 10 Healthy Cities in then member states of the European Union. The study design was selective and focused on a limited set of parameters mostly concerned with urban health policy networks. Additional reporting happened in Price and Tsouros (Price and Tsouros, 1996), a collection of case studies submitted by designated Healthy Cities throughout the phase. From Phase III, it was recognized that evaluations should be integral to the operations of the network, and that regular time series of data collection and analysis should be driven by a broad mixed-methods approach profoundly related to the aspirations of WHO and Healthy Cities themselves (De Leeuw, 2009). The proposed MARI (Monitoring, Accountability, Reporting and Impact) research framework was applied and findings were reported in Tsouros and Green (Tsouros and Green, 2009). An assessment of the efficacy of the research tools showed that, because of survey exhaustion and data overload, city representatives did not feel the exercise yielded substantive added value. For Phase IV, MARI was slimmed down to annual reporting templates (ARTs) and one final questionnaire in which key areas of Healthy City action were highlighted (De Leeuw, 2012; Tsouros and Green, 2013). A methodology started to crystallize in which research questions and approaches were defined and refined in collaboration with all relevant stakeholders, reflecting the naturalistic Fourth Generation Evaluation approach advocated by Lincoln and Guba (Lincoln and Guba, 1986).

Discussions about evaluating Phase V (2009–13) started early. It was felt by city coordinators and city lead politicians as well as WHO that the commitment to rigorously engage with evaluation efforts engrained in designation (for individual Healthy Cities) and accreditation (for Healthy City Networks) criteria required further operational, logistical, aspirational and conceptual development. The emerging methodological realm of realist evaluation and realist synthesis was proposed to stakeholders as a useful strategy to elicit the evidence that both local practitioners, politicians and principals as well as national and international leaders would require to further develop Healthy City approaches and validate and substantiate actions and values that work and make a difference.

A series of concept notes for such an evaluation approach was discussed with Healthy City representatives and advisory boards, WHO and academic peers. A programme logic (Figure 1) was iteratively developed adopting the concept and principles of Realist Evaluation which accounts for context, and the process of change from baseline to outcomes and impacts. This broad framework is applied to the Phase V Evaluation using a ‘realist synthesis’ approach to the collection and analysis of evidence (Pawson et al., 2004). This methodology is a better fit for our purpose than orthodox public health research paradigms. Figure 1 outlines stages and links between the elements of the European Healthy Cities process and how this can be illuminated by the evidence from the various research instruments we have used.

Realist review has emerged as a strategy for synthesizing evidence and focuses on providing explanations for why interventions may or may not work, in what contexts, how and in what circumstances. Traditional systematic reviews may provide evidence that certain interventions work, but not under which parameters. It is therefore important to seek stronger foundations in theoretical approaches to intervention development and implementation, as advocated, for instance, by Birckmayer and Weiss (Birckmayer and Weiss, 2000) in their Theory-Based Evaluation approach. The rigorous application of theory or conceptual framework would explain how things work and in what context (Clavier and de Leeuw, 2013).

Realism as a philosophical perspective on the scientific endeavour recognizes that even for complex issues and environments causal explanations are feasible; that the reality under study is largely constructed by its social actors; and that there is no situation in which the evaluators can divorce themselves from the social context in which those realities are constructed. Realist evaluation thus seeks to identify underlying causalities and explore how they work under what conditions. The ‘formula’ that describes this is expressed as ‘context + mechanism = outcome’ (Pawson and Tilley, 1997). Realist synthesis adds a dimension to realist evaluation; it inserts the notion that it may not be necessary to assess and evaluate chains of causality to demonstrate that interventions are effective (or not). The aim of a realist synthesis is ‘. . . to articulate underlying programme theories and then to interrogate the existing evidence to find out whether and where these theories are pertinent and productive. Primary research is examined for its contribution to the developing theory. . . ’. For instance, Langford
et al. (Langford et al., 2015) applied a realist synthesis gaze to a Cochrane Review on health promoting schools and explained that the existing evidence on the complexity of causal linkages driving the obesity epidemic, connected with the Cochrane Review schools evidence, suggested particular combinations of contexts and mechanisms to yield a reduction in chronic disease.

In the discussions with Healthy City stakeholders, it was therefore important to develop a conceptual model that distinguishes between context and mechanisms, in order to provide strong indications for likely outcomes. Our dynamic model shows a pathway linking prerequisites (leadership; vision and strategy; structures and processes; networks) through activities (policies; programmes; projects) which should ‘make a difference’ to city status in the areas of determinants of health, lifestyles and health outcomes. These elements all form part of the formal process through which interested local governments become designated as ‘Healthy Cities’. Whitfield et al. (Whitfield et al., 2013) first applied the realist synthesis approach to municipal investments by four member cities of WHO Healthy Cities European Network. This approach allows for detection of changes over time: the time dimension is explicitly included in the different research methods.

Although the arrows in the model (Figure 1) suggest linear and causal relationships between the three elements of the programme, earlier evaluation work (e.g. De Leeuw and Skovgaard, 2005) shows that these relationships are reciprocal and dynamic. It is therefore not always possible to distinguish between ‘dependent’ and ‘independent’ variables.

To be designated as a ‘Healthy City’ in each phase, city administrations take an official position and express commitment to attributes of local action for health such as: governance, equity, working in partnership, assuming leadership for health, engage all stakeholders and communities to participate in health and move towards sustainable policy development for health (Table 1, WHO, 2009). None of these attributes exist in isolation, and are found in different mixes and balances, stages of evolution and implementation and configurations in the designated Healthy Cities. The unique context of a Healthy City for development and implementation of actions for health (driven by these attributes) must be acknowledged. Thus, an evaluation approach that values and celebrates the unique contributions of each city has been developed.

In order to be designated, cities commit to a number of core criteria for Phase V (Table 2). These criteria were used

---

Fig. 1: Programme logic, Phase V healthy cities evaluation. GEQ, General Evaluation Questionnaire 2014; Case studies: thematic, strategic, and proudest achievement case studies 2009–13; ART, Annual Report Templates 2009–13; indicator profiles, quantitative data, city fact sheets.
to drive case study submissions as one of the key data sources for the evaluation.

**CASE STUDIES AS DATA SOURCES**

Case study research in itself is a powerful tool to generate insights with good validity and generalizability (Flyvbjerg, 2006). In the Phase V Evaluation exercise, the potential of this tool is enhanced by other methods and tools. The development of case studies is important for a number of reasons:

- they capture local stories for local conditions;
- they contribute to the establishment of continuity in corporate and community history;
- they allow for show-casing of unique and rewarding experiences;
- they can generate a breadth and depth of data and information that may otherwise not be captured.

A good case study for research purposes is not simply a story. It has a composition and direction that allows for the full capture of experiences and information. Three perspectives in generating such case studies have been deployed:

- structured case studies, where cities are requested to construct narratives in a highly structured manner around key attributes and visions of the programme;
- free-form case studies (with a pre-determined structure and word length) allowing cities to share their successes;
- narratives generated in communication sessions such as round-tables, workshops, etc.

The relatively 'open' evaluation architecture would allow for a much more realistic description of issues, challenges and solutions which cities and their institutions have come up with: opening the box enables cities to choose their own highlights.

---

**Table 1:** Phase V Evaluation logic framework matching unique Healthy Cities attributes to possible impact at different levels (WHO, 2009)

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Has adoption made a difference for strategic action at executive levels (e.g. Mayor’s Office, establishment of structures and accountabilities)</th>
<th>Has adoption made a difference for operational action with constituents (e.g. stakeholders, populations, industries)</th>
<th>Has adoption made a difference to the health (experience) of communities, groups, issues and health status?</th>
<th>Has adoption made a difference for the extent and intensity of communicating, linking to, and action on social determinants of health?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partnership</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Policy-making</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Summary core designation criteria for Phase V (WHO, 2009) (http://www.euro.who.int/_data/assets/pdf_file/0009/100989/E92260.pdf)

1. Health and health equity in all policies
2. Caring and supportive environments
   - Subtopics: Better outcomes for all children; age-friendly cities; migrants and social inclusion; active citizenship; health and social services; health literacy
3. Healthy living
   - Subtopics: Preventing non-communicable diseases; local health systems; tobacco-free cities; alcohol and drugs; active living; healthy food and diet; violence and injuries; healthy settings; well-being and happiness
4. Healthy urban environment and design
   - Subtopics: Healthy urban planning; housing and regeneration; healthy transport; climate change and public health emergencies; safety and security; exposure to noise and pollution; healthy urban design; creativity and liveability

---
Methodologists have argued that the generation and analysis of a large quantity of case study material would allow for rigour in the research process, for responsiveness to research and evaluation need of many stakeholders and for the generation of commonly accepted scholarly quality measures such as validity and reliability.

**DATA COLLECTION METHODS AND TOOLS**

The realist synthesis methodology for the Phase V Evaluation adopted a multi-method approach, building on trialled and tested approaches in earlier phases. This allows for triangulation of data and information.

The overall Phase V evaluation approach was validated with the Healthy City offices in Copenhagen and Belfast and with the Advisory Group (2012), and at the Izmir Business meeting (2013). Refinements in the case study tools were implemented, and workshops (with compulsory attendance of local coordinators) organized during the Izmir Business Meeting to pre-test and brief the logic and expectations for the evaluation and data collection methods.

The core of the approach consists of structured case studies around Healthy City attributes (colloquially called ‘strategic case studies’); case studies describing content areas as determined by the Phase V designation requirements (‘thematic case studies’); and ‘free form’ case studies that would allow cities to showcase their proudest achievements.

Cities were sent fillable PDF case study templates—both strategic and thematic (see Supplementary data). Case studies returned were coded using the qualitative software package NVivo, allowing for labelling, coding, connecting and assessing qualitative text-based data. A codebook was developed, peer-assessed and updated in the process grounded in the programme logic of Figure 1, grouping nodes of codes (Main Codes) in the different programmatic areas, with a series of more specific codes (Subcodes) for each of the main codes. The coding was performed by two independent coders with quality assessment to ensure consistency. We did not have an opportunity to calculate inter-observer $\kappa$ measures for reliability purposes as resourcing did not allow for a full coding and comparison exercise. However, a random selection of case studies was coded by both coders and compared, with inconsistencies being assessed by an independent evaluation expert, leading to full consistency of all codes. During the interrogation phase of the evaluation, in which assessors reviewed materials particular to their field of expertise, further refinements were added to the coding, without critical changes having to be implemented (cf. Bazeley and Jackson, 2013).

Following this inductive process and inter-observer quality control, the final codebook consisted of 54 Main Codes, with 33 Subcodes, and 7 of these contained detailed further specified codes.

The second data collection tool consisted of a General Evaluation Questionnaire (GEQ)—both for city level and for National Networks of Healthy Cities. The conceptual approach again followed the logic of Figure 1 and asked for self-assessment on prerequisites, activities and city status for 3 years: 2009 (Phase V start), 2014 (Phase V end) and informed guesstimates for 2019 (Phase VI end).

The GEQ was developed in consultation with (and pre-tested by) the Healthy Cities coordination team. GEQ consisted of 48 questions aligned with the strategic attributes as well as thematic areas cities committed to working on as part of their designation. For each of these areas, city coordinators were asked to rate the situation in their city (on a 10-point scale) at the start of Phase V, at its conclusion, and—in case they intended to continue in Phase VI (2014–19)—at the end of the next stage of their designated Healthy City status. This approach would capture in a condensed format progress that cities supposedly submitted regularly through the mandatory Annual Reporting exercise throughout the Phase.

GEQ also required the provision of information on political and organizational changes in the local Healthy City environment, the level of commitment and support (e.g. through Council statements and decisions) and resourcing (budget size, numbers of full-time paid staff, volunteer staff) and efforts at capacity-building focused at professional development and community agency.

The interactive GEQ was rolled out electronically through SurveyMonkey, reminders were sent to coordinators and some of them approached personally to urge response.

The case studies and GEQ were further complemented by compiling information from existing data:

(i) The ARTs are project management support exercises, filled in every year by designated cities. Filling out the ARTs is a designation requirement for all cities, and time series data from these have been compiled and are available to add detail and increase accuracy to the evaluation.

(ii) City Health Profiles (links or documents), containing local level health (social/environment/other) indicators have been collected from cities through the years and served as a reference for pre-/post-phase analysis where available and possible.

(iii) Cities’ initial documentation that leads to their entry into the project is kept in designation portfolios; these were cross-examined when appropriate to see initial
(baseline) city commitment and plans for activity in the specific Phase V themes and subthemes.

(iv) Healthy Cities Indicators. During Phase I (1986–2002), a set of 53 indicators was produced to assist cities in gathering appropriate data to describe health in their cities. Later some indicators were excluded, as the information they provided was not reliable or appropriate (Webster and Price, 1996; Webster and Sanderson, 2013). The wording of others was changed in order to define them more clearly. The result was a more concise set of 32 indicators.

(v) Yet, the currency of most indicators across cities in different countries is still challenged, and we embarked on an exercise to mine data from other databanks, including Eurostat (a major European database compiling data at various levels of governance, so-called ‘NUTS’ levels, Nomenclature des Unités Territoriales Statistiques, see, e.g. Stampach and Geryk, 2012). Eurostat NUTS2 data are available for all Healthy Cities in Europe, except for the Russian Federation, Israel, Serbia and Bosnia-Hercegovina. Alternative databases were sourced for these, but comparability of data could only be assured for Israel.

DATA RESPONSE AND QUALITY

City classifications

For the further analyses of our responses, we required a more fine-grained typology of designated cities. Information on assets, economic basis and demographic characteristics could be gleaned from the quantitative data sources, but it was considered essential to provide analysts with typologies of the 99 cities so that developmental patterns might be distinguished and differences attributed to categories. Distinguishing between contexts for different Healthy Cities would also allow for a more detailed response analysis, ascertaining that no one type city biased our analyses. Our categorizations became the following:

- Old–new. One of the assumptions over the lifetime of the European Healthy Cities Project has been that continued designation status over different phases would impact on the sophistication of the actions and policies developed by the Healthy City. In order to allow for testing this assumption, we provided data to analysts describing ‘New’ cities (those that were designated as a Healthy City for the first time in Phase V) and ‘Old’ cities (those that had a first designation any time during Phases I–IV).
- City size. There is no standard categorization of city size, or even a reliable distinction between a ‘village’, a ‘city’, or a ‘megacity’, although Doxiadis (Doxiadis, 1970) proposed a logical approach to city size evolution, eventually leading to the ‘ecumenopolis’ (a fully urbanised environment). In some urban planning literature, urban footprints with a city size between 1 and 5 million are labelled a ‘medium sized city’. We proposed three categories: small cities (with a population < 50 000), medium-sized cities (50 000–99 999) and large cities (larger than 100 000) as this gave three groups with roughly the same number of cities. However, data sheets were prepared for analysts with specific data for each city. This should allow analysts to distinguish by city size to their own need.
- Geo-political/governance groupings. It was felt that there are differences in the types of cities that could be attributed to, for instance, democratic tradition, legacy of governance approaches, ‘style’ of the market economy and perhaps even weather patterns. A number of classifications were weighed and assessed, with a consideration that the number of cities in each resultant category should ultimately allow for a degree of (statistical) analysis that would make sense. Geographical classifications are notoriously difficult to establish for many reasons, including historical and geopolitical legacies (Stampach and Geryk, 2012). We considered applying the commonly used United Nations Geoscheme (UN Statistics Division, 2013), but this did not seem to enable us to fully appreciate health system and governance differences. One approach that was considered looked at models for the delivery of healthcare, i.e. the Bismarck vs the Beveridge model (De Leeuw, 2013), but current European realities more often show the development of mixed systems. Another approach involved the application of Esping-Andersen’s (Esping-Andersen, 2013) welfare state typology, leading to Bambara’s (Bambara, 2007) four groups of a liberal (with as a dominant example the USA), conservative (Germany), social-democratic (Sweden) and Familistic/Confucian (Italy) nature. We also considered membership of the European Union and Council of Europe. We mixed and weighed these considerations, and ultimately quite arbitrarily, we grouped European countries into four categories: OECD (classic market capitalist welfare states); Mediterranean (classic market capitalist welfare states with a slightly instable political past); New Europe (nations having joined the European Union since 1989); and New Independent States (with a history of association with socialist states such as the USSR or the People’s Republic of China—in the case of Albania). The result is shown in Figure 2.

Responses

The total number of designated Healthy Cities in Phase V was 99. They entered the network continuously throughout
the Phase (2009–13), with a small number joining only in 2013 ([Tsouros et al., 2014], p. 3); these have not been considered likely to respond comprehensively to our data collection invitations. All cities were requested to submit at least one strategic, and one thematic case study. In total, 159 case studies were submitted. The city response rate was overwhelming: 73 cities submitted one or more case studies (resulting in an overall response rate of 74%). Not all cities, however, submitted both strategic and thematic case studies. Sixty-six (67%) submitted thematic case studies, and 68 (69%) submitted strategic case studies. Fourteen cities submitted more than the two required case studies, with two cities four, one city five and one city seven case studies equally distributed over thematic and strategic. In addition, six cities also sent free form case studies documenting their proudest achievements (Tables 3 and 4).

### General evaluation questionnaire

Seventy-one cities responded to the GEQ which yields an excellent response rate of 72%. The data supplied through SurveyMonkey were exported to SPSS and Windows Excel data files and made available to analysts. Some initial analyses and data integrity checks were performed at the evaluation coordination level.

### Annual reporting templates

Four ARTs had to be submitted for each of the first four years of Phase V. For 2013, ARTs questions were incorporated in the GEQ, so ART was not requested to be submitted (Table 5).

### Healthy cities indicators

Four categories with 32 indicators were searched for in Eurostat:

A. Health Indicators
B. Health Service Indicators
C. Environmental Indicators
D. Socioeconomic Indicators

Twelve Eurostat indicators were found to have data on local level for all cities/countries. Eurostat does not cover Russia, Bosnia and Herzegovina, Serbia and Israel. For all other cities, the following data were available:

- A1. Mortality
- C1. Atmospheric pollution

---

**Table 3:** Crude response numbers against pre-Phase V and Phase V-designated cities

<table>
<thead>
<tr>
<th></th>
<th>New (for the first time designated in Phase V)</th>
<th>Old (for the first time designated pre-Phase V)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designated</td>
<td>28</td>
<td>71</td>
<td>99</td>
</tr>
<tr>
<td>Case study response</td>
<td>18</td>
<td>55</td>
<td>73</td>
</tr>
<tr>
<td>(# cities)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case study response</td>
<td>43</td>
<td>116</td>
<td>159</td>
</tr>
<tr>
<td>(# cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEQ response</td>
<td>19</td>
<td>52</td>
<td>71</td>
</tr>
<tr>
<td>ART 2012 response</td>
<td>21</td>
<td>50</td>
<td>71</td>
</tr>
</tbody>
</table>

**Table 4:** Cities and case studies submission

<table>
<thead>
<tr>
<th></th>
<th>Number of case studies</th>
<th>Submitted by number of cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number received</td>
<td>159</td>
<td>73</td>
</tr>
<tr>
<td>Strategic</td>
<td>73</td>
<td>68</td>
</tr>
<tr>
<td>Thematic</td>
<td>80</td>
<td>65</td>
</tr>
<tr>
<td>Proudest</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**Table 5:** Summary of ARTs responses during Phase V

<table>
<thead>
<tr>
<th>Year</th>
<th>Response</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009/10 published June 2010</td>
<td>47 of 54 returned</td>
<td>87</td>
</tr>
<tr>
<td>2010/11 published June 2011</td>
<td>66 of 84 returned</td>
<td>79</td>
</tr>
<tr>
<td>2011/12 published June 2012</td>
<td>64 of 95 returned</td>
<td>67</td>
</tr>
<tr>
<td>2012 published June 2013</td>
<td>71 of 99 returned</td>
<td>72</td>
</tr>
</tbody>
</table>
DISCUSSION

Limitations of the evaluation approach
The very complexity of cities and their layers and domains of governance require a corresponding method to understand the context and impact of a healthy city approach. Context is important. The divergence of socio-political conditions and governance arrangements is challenging (Green, 1998; Kickbusch and Gleicher, 2012). Each local government area has a unique combination of characteristics, is located in a specific place and has profound roots in social, cultural and political history. Although the WHO Regional Office for Europe provides a common set of objectives, values and ideals for further development, the starting point and evolutionary pattern varies between cities and sometimes between communities and neighbourhoods within cities. The expansion of the European Union, for instance, has influenced opportunities for developing local government infrastructure. Austerity measures have had differential effects on cities across the European Union.

Good research attempts to identify confounding factors, that is, issues or events that disturb or complicate the dependent and independent variables that are the focus of the research. An example is the complicating factor of pollution from vehicular traffic, which diminishes the effects on health (dependent variable) of outdoor physical activity (independent variable). Another is the complicating factor of austerity, which may diminish the effects of signing up for the prerequisites of a healthy city. This in itself has been shown to be essential to embedding health across social and political sectors (De Leeuw and Skoggaard, 2003). The Phase V evaluation had limited capacity to identify and account for such factors. WHO is clearly not the only actor, and possibly a lesser agent in national contexts, driving health development. National governments are influential, limiting cities’ control over their funding, future and resilience. Within cities, governance arrangements allocate responsibilities and budgets between partner agencies.

This evaluation embraces the realist synthesis approach developed by Pawson et al. (Pawson et al., 2004), and recommends a continuation of the healthy dialogue with the policy community throughout and beyond the research process, from the initial expert framing of the problem to the final judgement on what works: ‘The tasks of identifying the review question and articulating key theories (of change) to be explored cannot meaningfully occur in the absence of input from practitioners and policy-makers’ (Pawson et al., 2005). A strong indicator of the importance of this position is the fact that, having engaged in this dialogue, our response rates can be considered excellent.

Limitations: non-response
A non-response analysis was carried out by the evaluation coordination team (for full response statistics, see above). Non-response and late-response cities were approached by e-mail and telephone to query what factors contributed to a less than diligent response:

(i) a few cities reported that the evaluation requests created a double/additional work burden, perceiving that the annual reporting exercise and/or designation process should have been sufficient to generate data;
(ii) political changes and commitment in some cities caused a lower prioritization of evaluation on the agendas of those cities;
(iii) structural changes in the location or appointment of Healthy City coordination offices;
(iv) the fact that the evaluation instruments were deployed in English only caused translation and interpretation issues in some cities;
(v) some cities were designated towards the end of Phase V, and some of these cities felt they were not in a position to report on progress yet;
(vi) some cities had chosen not to continue their Healthy City status in Phase VI and felt the evaluation exercise was not in their interest.

We subsequently performed analyses to see whether the non-response group displayed any patterns that would suggest any particular characteristic for (a group of) cities resulting in bias for our entire data set. No significant dissimilarities were found.

Having assured the quality and representativeness of the data, our data bases were offered to teams of analysts

C3. Percentage of water pollutants removed from total sewage produced
C4. Household waste collection quality index, total waste generating (both commercial and household, not separately listed)
C6. Relative surface area of green spaces in the city
C7. Public access to green space
C11. Cycling in city
C12. Public transport
C13. Public transport network cover
D3. Unemployment rate (data ranges between 2003 and 2012)
D8. Percentage of disabled persons employed (only available for 2002)
that report their findings in this Supplement. A 2-day plenary session with all 15 team members present kicked off the detailed interrogation of the enormous variety and diversity of data. Analysts were assisted by two data managers who could provide access to databases and provided detailed qualitative and quantitative data. However, analytical capacity in some cases was limited—we refer to individual papers in this Supplement to highlight particular challenges and limitations.

SUPPLEMENTARY DATA

Supplementary data are available at Health Promotion International online.

REFERENCES


United Nations Statistics Division. (2013) Composition of macro geographical (continental) regions, geographical sub-


