Recent Progress and Challenges in Assessing the Effectiveness

In recent years, air quality has improved in many locations around the world. In parallel, a growing number of studies are measuring whether predicted improvements in air quality and health outcomes resulting from specific interventions can, in fact, be detected, a field of research known as health outcomes or accountability research. Despite some early success stories, it remains challenging to link specific interventions to specific improvements in public health because good quality data are scarce, complex programs with multiple stage interventions affect air quality in different ways, and advanced statistical approaches are needed to address confounding issues—especially when evaluating regulations that are implemented over an extended period of time. In this article, we describe recent progress in evaluating the effectiveness of air quality regulations, and illustrate some of the challenges often encountered in this type of research.
There has been a long-standing interest in measuring the effectiveness of air quality regulations and other interventions to improve air quality, as indicated in the accompanying articles by Hubbell and Pope elsewhere in this issue. Over the past few decades, a body of evidence has started to emerge; some early studies were quite promising because they showed that interventions led to dramatic, sudden improvements in air quality and were associated with clear improvements in health outcomes. These included the well-known examples of a steel mill closure in the Utah Valley, a ban on coal sales in Dublin, Ireland, and regulations to reduce sulfur in fuel in Hong Kong.

In this article, we discuss recent insights from follow-on studies of the interventions in Dublin and Hong Kong that have shed further light on the original findings. We also revisit the evaluations of the 1996 Olympic Games in Atlanta and describe recent results from the 2008 Olympic Games in Beijing, which can be viewed as “experiments” in which interventions lasted only a limited period of time. The Beijing study and a recent study of a wood stove change-out program in Montana were unique in that they were conducted prospectively and followed a small number of participants before, during, and after the intervention.

In contrast, there have been relatively few studies of complex regulations that are implemented in multiple-year programs, as discussed by Pope. We describe a recent study of restrictions of power plant emissions that emphasizes the challenges in evaluating regional scale air quality changes. Table 1 provides an overview of the interventions and studies discussed in this article (all funded by HEI), highlighting their strengths and weaknesses. For more detailed information on the general challenges specific to air quality and health outcomes research, we refer the reader to a series of previous papers.

Revisiting Early Success Stories
The 2002 study by Clancy and colleagues of the 1990 ban on the marketing, sale, and distribution of coal in Dublin, Ireland, has been a flagship illustration of a significant decrease in concentrations of black smoke that was associated with a noted reduction in cardiovascular and respiratory mortality. A subsequent study by Dockery and colleagues focused on coal sales bans in smaller Irish towns from 1995 to 2000 and compared the changes with observations in rural areas without a ban. In addition, the original analysis of mortality data in Dublin was expanded to include hospital admissions data. Dockery and colleagues report decreases (ranging from 45–70%) in black smoke in each of the Irish cities after their respective bans, but no clear pattern in sulfur dioxide (SO2) concentrations. The bans were associated with significant reductions in respiratory mortality and cardiovascular hospital admissions, but not cardiovascular or total mortality.

The ban on coal sales appears to have been associated with some health benefits, although the extent of those benefits in Dublin and other parts of Ireland was less than originally reported. Possible explanations are that the population outside of Dublin is relatively small, reducing the power to detect significant changes in mortality. In addition, the subsequent bans resulted in a lesser improvement to air quality because pollutant concentrations were lower at the subsequent starting points. A general downward trend in air pollutant concentrations over time and long-term declines in mortality rates may also have contributed. Finally, neighboring counties may have been affected indirectly due to economic factors (e.g., the price and availability of coal) and other trends that resulted from bans elsewhere (e.g., a general trend to replace coal furnaces with gas stoves).

A 2002 study by Hedley and colleagues focused on a 1990 reduction of sulfur in fuel in Hong.
Kong, and showed that SO₂ concentrations decreased by 45% during the five years after the intervention; particulate matter (PM) and nitrogen dioxide (NO₂) concentrations were not affected, but ozone (O₃) levels increased by 28%. Hedley and colleagues noted a substantial reduction in mortality over five years after the intervention and reported that a greater decline in mortality was noted in areas with a higher reduction in SO₂ concentrations.

In a subsequent study, Wong and colleagues evaluated the effects of subsequent regulations on mortality and life expectancy, with the additional goals of assessing long-term benefits of improved air quality over a 20-year period and looking at PM composition. They report decreases in several PM components, of which nickel and vanadium were the most consistent. Although the authors found an association between PM and PM-associated metals and mortality, they were unable to definitively link the intervention to changes in mortality. Possible explanations are that PM components are

### Table 1. Overview of Recent Evaluations of Air Quality Interventions Funded by HEI.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Study Design/Period</th>
<th>Strengths and Challenges</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ban on sale of coal in Dublin (1990) and other cities in Ireland (1995, 1998, and 2000)</td>
<td>Retrospective time-series; 1981–2004</td>
<td>Included non-ban control areas. Air quality improvement was less after subsequent bans in smaller towns; confounding by long-term air mortality and economic trends; small population outside Dublin</td>
<td>Clancy et al. (original study)⁴ Dockery et al. in press (follow-up)¹⁰</td>
</tr>
<tr>
<td>1990 sulfur restrictions in fuel in Hong Kong, China</td>
<td>Retrospective time-series; 1985–1995</td>
<td>Evaluation of PM components and statistical approach to assess life expectancy. Concurrent changes in Ni and V; no clear association of mortality with any PM component; limited success in assessing life expectancy changes</td>
<td>Hedley et al. (original study)⁵ Wong et al. (follow-up)¹¹</td>
</tr>
<tr>
<td>Traffic restrictions during 1996 Olympic Games in Atlanta, Georgia</td>
<td>Retrospective time-series; 1995–2004</td>
<td>Evaluated seasonal changes in surrounding years and included Southeast United States. Regional ozone improvement casts doubt on it being related to the traffic intervention.</td>
<td>Friedman et al. (original study)⁹ Peel et al. (follow-up)¹³</td>
</tr>
<tr>
<td>Industrial emissions and traffic restrictions during 2008 Olympic Games in Beijing, China</td>
<td>Prospective panel study; before (June 10–July 6), during (August 4–29), and after (October 6–31)</td>
<td>Substantial air quality improvement but difficulty in tracing this to specific interventions. Favorable weather also played a role. Unclear clinical significance of changes in biomarkers in healthy subjects.</td>
<td>Rich et al.²⁰ Zhang et al. in press²¹</td>
</tr>
<tr>
<td>2003 congestion charging scheme in London</td>
<td>Emissions modelling and air quality evaluation; 2001–2005</td>
<td>Small target area (inner city) and regulation not targeted at air quality per se; confounding by regional background pollution</td>
<td>Kelly et al.¹⁴</td>
</tr>
<tr>
<td>Wood stove change-out in Libby, Montana during 2005–2007</td>
<td>Prospective study in school age children, parent questionnaire; 2003–2009</td>
<td>Successful change-out program led to improved ambient air quality but indoor air was not always improved. Overall respiratory health benefits observed regardless of whether children lived in a home with a changed woodstove.</td>
<td>Noonan et al.¹⁶–¹⁸</td>
</tr>
<tr>
<td>2008 low emission zone in London</td>
<td>Collection of baseline air quality and health data before LEZ went into effect</td>
<td>Additional road-side monitoring. LEZ implementation is taking place in stages over multiple years; air quality improvement expected to be fairly gradual.</td>
<td>Kelly et al.²⁵</td>
</tr>
<tr>
<td>1990 Clean Air Act Amendments to restrict power plant emissions of NOₓ and SOₓ in the Eastern United States</td>
<td>Statistical analyses linking emissions and source-receptor data; 1999–2005</td>
<td>New statistical approach. Confirmed decreased PM₂.⁵ concentrations but uncertainty in models due to large number of variables and missing data.</td>
<td>Morgenstern et al. in press²⁶</td>
</tr>
</tbody>
</table>

Note: Several studies are follow-on research of earlier work, as indicated. Studies are listed in the order in which they are discussed in the text.
generally measured only once every six days, and the considerable spatial variability in air quality changes was difficult to account for in the statistical analyses. The authors estimated effects of the sulfur reduction on life expectancy based on analysis of daily time-series of PM and mortality; this approach met with limited success, due to difficulty in controlling for long-term trends in potential confounding factors.

Friedman and colleagues\textsuperscript{12} studied the traffic intervention in Atlanta, GA, during the 1996 Olympic Games. Results showed a decrease in acute care visits for pediatric asthma during the Games compared with the weeks before and after. This study also showed an associated decrease in O\textsubscript{3}, PM\textsubscript{10}, and carbon monoxide (CO) concentrations. However, it remained unclear to what extent normal seasonal patterns in pollutant concentrations or health outcomes may have influenced the results.

In a subsequent study, Peel and colleagues\textsuperscript{13} confirmed that O\textsubscript{3} concentrations were approximately 30\% lower during the Olympic Games; PM\textsubscript{10}, CO, and NO\textsubscript{2} concentrations were also slightly lower. However, they showed that there were similar reductions in O\textsubscript{3} concentrations in several other cities throughout the Southeast United States that were not affected by the traffic intervention. In addition, Peel and colleagues observed little or no evidence of overall reductions in traffic or reduced emergency department visits for respiratory or cardiovascular health outcomes in adults or children during the Olympic Games when they took into account seasonal health trends in previous and following years.

The follow-on study by Peel thus illustrated the crucial importance of using control areas, evaluating appropriate time windows surrounding interventions, and properly adjusting for seasonal and other trends that may influence the results. Ultimately, it called into question whether the findings of the original Friedman study could be attributed to the traffic restrictions and whether the traffic actions taken could have been designed more effectively.

Other studies have noted similar difficulties in finding substantive air quality changes associated with traffic interventions, in part, due to the limited area covered by the intervention (i.e., city center only), limited road side monitoring that might have been able to show more definitive air quality improvements, and confounding by regional weather patterns.\textsuperscript{14} Potential benefits in terms of improved mortality were estimated to be low.\textsuperscript{15}

**Prospective Studies of Local Interventions**

When there is sufficient lead time before a particular regulation goes into effect, unique opportunities exist to conduct prospective evaluations, for example, panel studies that follow a small population before, during, and after the intervention and actively collect exposure and health information. One such intervention was the change-out of wood stoves in a rural mountain community.\textsuperscript{16-18} Noonan and colleagues found that the wood stove changeout was effective in reducing ambient pollutant concentrations, but that indoor concentrations did not always improve as expected. Possible reasons include variability in stove operation, presence of other indoor sources, and the fact that
Clean wood stoves are still relatively high emitters compared to stoves using other fuels, such as natural gas. Children’s respiratory health was somewhat improved, although there were no differences among children from homes where stoves were changed compared to homes where stoves were not changed.

During the 2008 Olympic Games in Beijing, large-scale interventions took place, including the shutting down of factories and other high-emitting stationary sources in the region, as well as restrictions on high-emitting vehicles and car usage in general. A number of recent papers have described the air quality changes, which included 13–60% reductions of PM and gaseous pollutants, except for an increase in O3 concentrations of approximately 20%. However, while the range of interventions appears to have contributed importantly to a reduction in air pollution, it remains difficult to establish a cause-and-effect relationship between specific actions and the observed air quality changes given the many actions taken and their staged implementation (e.g., some measures to curb industrial pollution started the year before). Favorable weather conditions during the Games also played an important role in reducing air pollutant levels. Nonetheless, the substantial air quality improvement during the Games—in which the emissions reductions from industrial sources and traffic played a major role, and which was accompanied with a worsening after the interventions ended—provided an excellent opportunity for health outcomes studies.

A prospective study led by Zhang and colleagues followed approximately 130 healthy medical students and measured cardiovascular biomarkers before, during, and after the Beijing Olympic Games. They observed small improvements in several cardiovascular measures, although a few changes were in the opposite direction from what was expected. The authors concluded that “changes in air pollution levels during the Beijing Olympics were associated with acute changes in biomarkers of inflammation and thrombosis and measures of cardiovascular physiology in healthy young persons. These findings are of uncertain clinical significance.” Another study followed a panel of school children for two years and reported reduced exhaled NO levels in breath during the Games; in addition, researchers reported improved heart rate variability in a small panel of taxi drivers and a reduction in outpatient visits for asthma in adults during the Olympic Games compared to before or after.

A prospective study in London assembled baseline information before a low emission zone (LEZ) was implemented, starting in 2008. This program included stages in which, starting with the most polluting trucks, polluting vehicles were banned from entering Greater London, or would pay large fines if they did. The study added road-side monitoring for better tracking of traffic-related air quality changes, and collected PM on filters for evaluation of oxidative stress potential. The investigators also developed approaches to obtain health outcomes data from general practices (rather than mortality and hospital admissions data) and improve exposure assessment by including traffic emissions and distance to road segments in their modeling. Given that the last stages of the LEZ implementation were to occur in 2012, an assessment of the effectiveness has not yet begun.

Progress in Evaluating Long-Term Regulations

As discussed in the accompanying Pope article, evaluations of long-term air quality changes are quite complex. Morgenstern and colleagues recently sought to evaluate the effect of national limits on NOx and SO2 emissions from power plants implemented under Title IV of the 1990 U.S. Clean Air Act Amendments to address the growing acid rain problem. They used a statistical analysis linking emissions and air quality monitoring...
data to predict PM$_{2.5}$ concentrations at different distances from air quality monitors before and after the installation of emissions-control equipment on the targeted stationary sources. Thus, the contributions of emissions from specific sources to air quality were assessed at distances as close as 0 to 12.5 miles (locally) and as far away as 400 to 500 miles (regionally). (An evaluation of health effects was not included at this time.) They found that the policy was associated with an average reduction of 1.07 (± 0.11) µg/m³ at PM$_{2.5}$ monitors in the eastern United States during the study period, which was similar to projections generated through the Community Multiscale Air Quality Model. It should be noted that this data-driven source–receptor approach is potentially useful, but that there are reservations about potential uncertainty in the results owing to the large number of variables included in the models and challenges in estimating concentrations due to a fairly substantial amount of missing monitoring data.

Conclusions
The first generation of outcomes studies provided much valuable information. Although some initial encouraging results could not be fully replicated in more detailed follow-on studies, new prospective studies have provided evidence of improved health in small populations that were monitored closely. It is important that the lessons learned be translated into guidance for the design and statistical analysis of studies of air quality interventions. As described in this article, many challenges can hamper a study’s ability to detect effects, including limitations in the availability of detailed monitoring data (e.g., few road side monitors that would be useful to evaluate traffic interventions), the influence of regional weather patterns, a lack of study power due to small population size, a lack of appropriate control populations, and other factors. At the same time, even studies that did not find expected effects can and have provided guidance about design elements of the interventions (e.g., the nature and extent of traffic actions) that might be re-visited in future interventions to improve their effect.

At this stage, there is clearly a need for additional studies that can take advantage of the lessons learned so far by carefully designing approaches to tackle these continuing challenges. Common study design issues that remain important are the selection of appropriate control populations, assessing what would have happened to air quality if the regulations were not implemented or different regulations were implemented (so-called counterfactual approaches to handle weather confounding; Air quality regulations of interest have been identified but linkage will be challenging.

Common study design issues that remain important are the selection of appropriate control populations, assessing what would have happened to air quality if the regulations were not implemented or different regulations were implemented (so-called counterfactual approaches to handle weather confounding; Air quality regulations of interest have been identified but linkage will be challenging.

Table 2. New HEI-Funded Health Outcomes Studies of Long-Term Regulatory Efforts at the Regional and National Scale.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Study Design/Period</th>
<th>Strengths and Challenges</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>California air quality control programs in general</td>
<td>Children’s Health Study cohorts; 1993–2012 (includes 3 cohorts)</td>
<td>Existing cohort study is providing high quality health data over two decades; potential to study low versus high exposure areas and socioeconomic status; Complex sets of regulations; link to specific air quality regulations has not yet been made</td>
<td>Gilliland et al. (conference abstract)27</td>
</tr>
<tr>
<td>2006 California Emission Reduction Plan for Ports and Goods Movement</td>
<td>Air quality modelling; 2 year period before and after implementation in 2006</td>
<td>Evaluating goods movement corridors versus other transportation corridors and non-transportation control areas; the “Plan” includes many small measures (e.g., related to cleaning up specific fleets); 2006 implementation was gradual and some measures are still being implemented</td>
<td>Meng et al. (conference abstract)28</td>
</tr>
<tr>
<td>Effect of Clean Air Interstate Rule and heavy duty diesel and low sulfur rules in Atlanta</td>
<td>Emissions and air quality modelling linked to time-series approach; 1993–2012</td>
<td>Strong continuous air quality and health datasets; exploring approaches to handle weather confounding; Air quality regulations of interest have been identified but linkage will be challenging</td>
<td>Russell et al. (conference abstract)29</td>
</tr>
<tr>
<td>PM NAAQS and related state implementation plans (SIPs)</td>
<td>Statistical methods development</td>
<td>New approaches and focus on attainment status; SIPs may also affect air quality in neighboring counties</td>
<td>Zigler et al. (conference abstract)30</td>
</tr>
</tbody>
</table>
air quality scenarios), selecting time periods surrounding the implementation of the regulatory action, as well as appropriately adjusting for unmeasured and potentially confounding risk factors and regional meteorological effects. Continued exploration of alternative approaches and additional methods development, especially for evaluating complex long-term regulatory actions and accounting for the individual and combined effects of multiple pollutants, is recommended. Currently, HEI is funding four new studies of long-term national and regional scale regulatory efforts (described in Table 2) that are expected to provide further insights in the near future.

References