Predicting Housing Abandonment with the Philadelphia Neighborhood Information System

Amy E. Hillier* Dennis P. Culhane†
Tony E. Smith‡ C. Dana Tomlin**

*University of Pennsylvania, ahillier@design.upenn.edu
†University of Pennsylvania, culhane@upenn.edu
‡University of Pennsylvania, tesmith@seas.upenn.edu
**University of Pennsylvania, tomlin.dana@verizon.net

PREDICTING HOUSING ABANDONMENT WITH THE PHILADELPHIA NEIGHBORHOOD INFORMATION SYSTEM

AMY E. HILLIER
University of Pennsylvania

DENNIS P. CULHANE
University of Pennsylvania

TONY E. SMITH
University of Pennsylvania

C. DANA TOMLIN
University of Pennsylvania

ABSTRACT: *Several large US cities, including Chicago, Los Angeles, New York, and Philadelphia, have developed information systems to distribute property-level housing data to community organizations and municipal agencies. These early warning systems are also intended to predict which properties are at greatest risk of abandonment, but they have rarely used statistical modeling to support such forecasts. This study used logistic regression to analyze data from the Philadelphia Neighborhood Information System in order to determine which properties were most likely to become imminently dangerous. Several different characteristics of the property, including whether it was vacant, had outstanding housing code violations, and tax arrearages as well as characteristics of nearby properties were identified as significant predictors. Challenges common to the development of early warning systems—including integrating administrative data, defining abandonment, and modeling temporal and spatial data—are discussed along with policy implications for cities like Philadelphia that have thousands of vacant and abandoned properties.*

Over the past 30 years, community groups, municipal governments, and researchers have identified early warning systems as potentially important tools for understanding, preventing, and managing housing abandonment. By using management information

Direct correspondence to: Amy E. Hillier, Cartographic Modeling Lab University of Pennsylvania, 210 S. 34th Street, Philadelphia, PA 19104. E-mail: ahillier@ssw.upenn.edu

JOURNAL OF URBAN AFFAIRS, Volume 25, Number 1, pages 91–105.

ISSN: 0735-2166.

systems (MIS) often in combination with Geographic Information Systems (GIS) technology to integrate property-level and aggregate data, these systems are intended to forecast which properties are at greatest risk of abandonment in order to inform planning, intervention, and research. Los Angeles (Community Information Technology Center, Neighborhood Knowledge Los Angeles) and Chicago (Center for Neighborhood Technology, Neighborhood Early Warning System) boast the largest and most accessible systems, but examples have also been found in Seattle (Bell & Kelso, 1986), Akron (Costa & Hanten, 1988), Minneapolis (Mardock, 1998; Miller, 1999), New York (Scafidi, Schill, Wachter, & Culhane, 1998), and St. Paul (Brown, 1999; Myott, 1999; Villares, 1999).

This article serves to introduce the Philadelphia Neighborhood Information System (NIS), an on-going collaboration between the University of Pennsylvania and the city of Philadelphia designed to integrate housing information into web-accessible mapping applications and to support early warning research. In addition to reporting the results of our predictive model for housing abandonment, this article describes some of the major challenges common to the development of such early warning systems: defining housing abandonment, integrating secondary data from multiple sources, incorporating temporal and spatial aspects of these data, and measuring the ability of various data elements to predict abandonment.

Though not all housing-related early warning systems have focused on housing abandonment, these systems are generally aimed at identifying properties at risk of physical decline. They have alternatively referred to deteriorating (Neighborhood Knowledge Los Angeles), distressed (Myott, 1999), vacant (Brown, 1999), and problem (Villares, 1999) properties. For all of them, the challenging task of conceptually defining housing decline is further complicated by the limits that data availability place on operational definitions. Early warning systems are premised on the idea that many different property and neighborhood characteristics indicate abandonment. As a result, they draw on secondary data from various municipal agencies and private vendors. Developers face the common challenge of integrating these data into user-friendly formats in the face of funding limitations, concerns about data quality and confidentiality, and sometimes contested inter-agency relations.

Because the promise of predicting future trends and events is built into these systems, they must also take temporal ordering into consideration. Data elements are only helpful when they are up-to-date, and data models are meaningful only to the extent that they reflect possible causal processes. The temporal nature of the data necessitates updating, date-stamping, and archiving, all of which demand well-defined data acquisition and storage strategies and the use of time-sensitive modeling techniques.

The spatial nature of the data complicates analysis as well. Common geography provides the framework for data integration and allows for simple visual representations through user-defined maps. But the tendency of abandoned properties to cluster within distressed areas raises questions about the scale of these spatial relationships and necessitates the use of space-sensitive statistical modeling techniques.

Finally, these systems need to assess the predictive value of the individual data elements they integrate. Only if tax arrearages, code violations, utility terminations, and housing sales actually correlate with housing decline should they be presented as indicators of risk. Assessing these relationships can be costly, but this step is crucial if these systems are to be used successfully for forecasting.

THE PHILADELPHIA NEIGHBORHOOD INFORMATION SYSTEM

The Philadelphia NIS supports municipal government and community agencies involved in housing services, neighborhood planning, and redevelopment by distributing integrated, updated, and map-enabled housing and neighborhood data through two different web-accessible applications. Eight different municipal agencies contribute parcel-level data at regular intervals, generally every six months. The ParcelBase, a password protected application that allows authorized users to identify and map individual properties based on the most current information in the NIS database, is accessible to registered staff from city agencies, community development corporations, and other community-based agencies that have contracts with the city. Access to the ParcelBase is restricted because some of the housing data is of a sensitive nature and because the license to use the city's parcel coverage in the application requires that users be registered through the city's Office of Housing and Community Development. The NeighborhoodBase is available to the public and enables users to query and map aggregate data and analyze trends. The NIS also generates research and policy reports based on analysis of this expanding data archive. The NIS is funded by the William Penn Foundation, Pew Charitable Trusts, and the University of Pennsylvania.

Defining Abandonment

By focusing on the distribution, rather than interpretation, of secondary data, most early warning systems have effectively sidestepped the issue of defining housing abandonment. By presenting a variety of raw secondary data elements as possible indicators of housing distress and abandonment, these systems can simultaneously serve a variety of constituencies. Community development corporations (CDCs) purchasing and renovating individual properties, community groups trying to hold owners accountable for property conditions, or municipal agencies planning large-scale redevelopment can all benefit from the system. Research applications of these systems require clear conceptual and operational definitions to measure and predict housing abandonment.

To abandon a house is to neglect the responsibilities of ownership regarding minimum functional, financial, and physical upkeep (Sternlieb, Hughes, Bleakly, & Listokin, 1974a). Abandonment is different from vacancy (Keenan, Lowe, & Spencer, 1999), a status referring to whether a property is occupied or not. Vacancy can be the result of normal turnover and can be temporary, long-term, or permanent. Knowing that a property is vacant indicates nothing about its financial or physical condition. Abandonment, on the other hand, is a staged process (Sternlieb & Burchell, 1973). Not all properties pass through the same stages, however, or in the same order as there are multiple causal pathways leading to abandonment (Sternlieb, Burchell, Hughes, & James, 1974b; Wilson, Margulis, & Ketchum, 1994). The abandonment process is better understood as a cycle rather than a linear series of events, i.e., even demolition of a residential building can lead to new investment. Furthermore, abandonment is often reversible because owners can resume meeting their responsibilities for upkeep, taxes, and utilities.

As suggested above, housing abandonment has three distinct aspects: functional, financial, and physical (Costa & Hanten, 1988; O'Flaherty, 1993). Functional abandonment occurs when a property is no longer fulfilling its role as a residence. Vacancy, suspension of mail service, utility terminations, and the sealing of doors and windows are all indicators of functional abandonment. Financial abandonment relates to disinvestment and occurs when owners stop meeting their minimum financial responsibilities and properties begin accumulating debt in excess of equity. Property tax arrearages, defaulted mortgages, and liens are all indicators of financial abandonment. Physical abandonment occurs when owners neglect the interior or exterior upkeep of a property. This could include minor violations of the housing code that relate to the safety and comfort of occupants or more serious structural problems. While it is often unsafe for people to live in these structures, vacancy is not a necessary precondition of physical abandonment, particularly since squatters or people with limited housing options may view these as viable sources of shelter.

Housing abandonment is not inevitable. It can be explained in part by population shifts and the natural aging and obsolescence of residential structures, but it is quite often undesirable and preventable (Galster, 1987). As Grigsby, Bratz, Galster, and Maclennan (1987) noted, "a continuing downward path of deterioration, commencing at the moment of first occupancy or a few decades thereafter, is not a structure's inevitable destiny" (p. 44). Abandoning a house involves a series of decisions, including the decision not to act (Sternlieb, et al., 1974b). The owner is generally responsible for the decision to abandon a property, but a host of other actors can also be involved. Municipalities can contribute to the abandonment process by failing to enforce the housing code, not sealing and demolishing properties, not collecting back taxes, or not financing redevelopment in certain areas. Police can influence the process by failing to enforce criminal or civil codes consistently. Lenders can affect property outcomes by refusing to make loans, realtors by steering prospective homeowners to other areas, journalists by not reporting on events, and neighbors by either moving away or simply withdrawing from neighborhood life.

Early warning systems are premised on the belief that there are multiple warning signs and opportunities for intervention throughout the process of abandonment. Owners receive code violations and complaints from tenants; city agencies receive complaints from tenants and neighbors; utility companies have unpaid accounts. These warning signs are not necessarily causal factors, but together they signal where a property is in the abandonment process.

While recognizing abandonment as a process is important in conceptual terms, it can make abandonment much more difficult to measure. As Sternlieb, et al. (1974a) explained, the key decisions about upkeep and investment "are obviously not publicly announced to municipal authorities. Thus, we cannot measure abandonment by monitoring these decisions directly, but only by observing the consequences of these actions" (p. 38). Almost inevitably, a publicly recorded housing-related event or condition must be chosen as a proxy with the understanding that it represents just one point in time and one aspect of abandonment. Much of the previous research about housing abandonment has used events relating to a building's financial situation as a proxy for abandonment (Arsen, 1992; Bender, 1979; Scafidi et al., 1998; White, 1986). Johnson and Accordino (2000), on the other hand, focused on functional abandonment and defined properties with long-term vacancies as abandoned. An alternative to both of these approaches is to view indicators of financial and functional abandonment as predictors of physical abandonment. While occupied properties where all financial obligations are being met can still experience physical abandonment, it is much more likely that these financial and functional indicators will precede physical abandonment. Furthermore, properties that have been completely functionally or financially abandoned can be returned to the housing stock, but this is less likely for properties that have been physically abandoned.

This research operationally defines housing abandonment as the point at which a property is declared to be imminently dangerous. Philadelphia's Department of Licenses and Inspections (L&I) judges that a structure is imminently dangerous when it presents a safety hazard to people and nearby structures and when the agency is prepared to begin the demolition process. This may include a court order that an owner start demolition within a fixed time period, although generally L&I begins the process of securing a contractor on its own and bills the owner for demolition-related expenses. Imminently dangerous properties are not always vacant, although L&I does post a sign on such properties prohibiting entry. Imminently dangerous properties were chosen rather than demolished properties because the city's ability to secure a contractor and begin demolition involves considerable financial, administrative, and legal considerations beyond an assessment of the buildings condition. On the other hand, deciding which properties to deem imminently dangerous is a more objective and systematic process. By focusing on a late stage in the process of physical abandonment, indicators of less serious physical distress can also be used as predictors of which properties may become imminently dangerous in the future.

Integrating Data

By design, early warning systems capitalize on the wealth of data collected by municipal agencies for administrative purposes, including billing and service delivery. Because agencies continuously update these data and generally collect them at the address level, they offer a valuable alternative or supplement to decennial US Census data. But because they are collected for non-research purposes by agencies with varying data collection and management procedures, the integration of these data can present a number of challenges. First, developers of early warning systems must secure data sharing agreements. Agencies contributing data to the NIS were asked to sign letters of authorization indicating the groups of users (city agencies, community organizations, researchers, the public) who could have access to their data at specific geographic levels (parcel, block, block group, census tract). The matrix is illustrated in Table 1.

As Sawicki and Flynn (1996) pointed out, integrating secondary data also raises concerns about data reliability and validity, as both systematic and random errors in the agency data can easily affect the early warning system. In the NIS, some errors are caught during the systematic data reviews conducted as data are received, but NIS users have provided an additional layer of quality control. Familiarity with their own data (for city agencies) and with particular properties and neighborhoods (for community organizations) enables them to see and report errors that are not obvious during the initial reviews.

The NIS includes more than 100 variables that were identified through a series of meetings and focus group sessions conducted by NIS project staff with personnel from city agencies and community organizations. Some agencies were unwilling or unable to

TABLE 1

Data Access Matrix

Attribute	City Agencies	Community Organizations	Public
Clean/Seal Date			
Type Clean/Seal Action			
Number of Actions			
Administrative Violations			
Building Violations			
Electrical Violations			
Fire Violations			
Housing Violations			
Plumbing Violations			
Zoning Violations			
Total Open Violations			
Inspection Date			

contribute requested data, some of the requested data was not available in digital format, and other data were inaccessible because of inherent limitations in extracting data from their stand-alone databases. Nearly all of the NIS variables cover the entire city, but at least one data set was based on a foot survey that was carried out in only part of the city. The housing attributes and digital photographs generated through this foot survey were integrated along with the citywide data, and the data elements are simply not displayed in the ParcelBase application for properties for which data are unavailable. Consistent with other early warning systems, the content of the NIS is driven by a combination of demand for data and data availability.

The NIS integrates data across agencies using the tencode, a nearly unique property identifier made up of a five-digit street code and five-digit house number. In the ParcelBase, characteristics of the properties are grouped according to their source agency using eight different tabs on a screen that pops up when a parcel is selected. A parcel coverage from the City Planning Commission serves as the geographic structure for organizing parcel-level attribute data from the various agencies. The application uses Microsoft Visual Basic and ESRI's MapObjects Internet Map Server to generate maps and tables in response to user queries. All of the data and the ParcelBase application itself reside behind two firewalls.

Temporal Relationships

Nearly all of the data elements included in the NIS and other early warning systems are time sensitive. Utilities that have been shut off can be turned back on; structurally compromised buildings can be demolished; residents can move in and out. This makes it essential that early warning systems involve regular data updates. Agencies providing data for the NIS are asked to provide updates every six months. All the datasets are maintained in a data archive. Only the most current data are displayed in the ParcelBase application, but the NeighborhoodBase allows users to identify the time period of interest. Ideally, all data elements should include a date stamp, indicating when a certain event (such as a utility termination) took place within the six-month period. But when this is not possible, it is critical that the acquisition date for a particular data set be displayed. The fact that municipal agencies change the content and format of their own data over time, even if the change represents an improvement, creates additional challenges for data integration and analysis of change over time. But early warning systems need to be flexible enough to adapt to such changes.

The temporal nature of the data is also central to building a predictive statistical model. Only if certain housing and neighborhood conditions or events precede abandonment can they be considered predictors. Ideally, a statistical model would indicate the amount of time by which an indicator precedes abandonment in order to address questions of temporal scale. How much of a warning does a particular indicator provide? Is abandonment imminent or several years off? Such modeling requires an extensive data archive with data reaching back 10 years or more.

For the model described here, properties were considered abandoned if they were designated as imminently dangerous between January and November 2000. (The Department of Licenses and Inspections provided the date on which the agency sent a letter to a property owner indicating that the structure was being declared imminently dangerous. The conditions that led to the declaration may have existed for some period of time before the letter was sent.) Properties that were declared imminently dangerous or demolished before January 1, 2000 were excluded from the analysis. To be considered as a possible indicator, variables had to reflect conditions or events that occurred or existed prior to January 2000, regardless of how far in advance they occurred. Some extended back as much as 20 years, while most reflected more recent events and conditions. As the NIS data archives expand, it may be possible to forecast longer-term abandonment trends, but at this point, the data only permits predictions of what properties will become abandoned over the next year or two.

Spatial Relationships

Early warning systems use common geographic references to integrate attributes of properties or neighborhoods, but they use mapping in different ways. Chicago's Neighborhood Early Warning System (NEWS) lists property attributes in response to user queries without locating the property on a map. Neighborhood Knowledge Los Angeles (NKLA), on the other hand, indicates the location of the requested address with a point along the street centerline in addition to listing the property's attributes. The Philadelphia NIS highlights the actual parcel on a parcel map, showing the location of the property in relation to the city or a particular block. It also enables users to identify nearby properties and their attributes. This has practical value for community developers, for example, who may need to work with owners of adjacent properties to acquire or renovate a property.

Beyond the practical value of increasing users' understanding of the geographic context of a particular property and facilitating spatial queries, mapping abandoned properties has important conceptual value. Housing abandonment may well relate to conditions in a neighborhood or regional area as much as the attributes of a specific property (Accordino & Johnson, 2000; Sternlieb, et al., 1974b). This interdependence is described in many different ways: externalities, spillover, and contagion effects (i.e., when housing abandonment impacts the surrounding area), and ecological, environmental, area, contextual, or neighborhood effects (i.e., when nearby conditions impact a property's likelihood of abandonment) (Kiefer, 1980; Odland & Balzer, 1978; Odland & Barff, 1982; Philippi, 1977; Wilson et al., 1994). The effect of housing abandonment on the likelihood of nearby abandonment has important implications for statistical modeling, raising concerns about spatial autocorrelation. The effect of conditions in an area on housing abandonment, on the other hand, presents challenges for determining the scale of the spatial relationships.

The presence of spatial autocorrelation, a likely problem in research when similar values on the dependent variable cluster together, can make estimates of the magnitude and, in particular, the significance of relationships between independent and dependent values unreliable (Bailey & Gattrell, 1995; Cressie, 1993). Spatial autocorrelation occurs when residuals from a regression model are similar for nearby observations, violating the assumption that observations are independent. This research successfully avoided problems with spatial autocorrelation by taking repeated random samples of 1000 from the original data set of 17,567 residential properties. These smaller samples included fewer nearby properties, thus reducing the amount of spatial dependence. The models in this study that included area-level variables (such as the concentration of demolitions within a quarter mile of a property) further decreased the significance of spatial autocorrelation.

Incorporating the effects of local or neighborhood conditions into the statistical model also raises questions about what is local. If the amount of nearby vacant housing correlates with housing abandonment, what is the scale of that spatial relationship? Do vacant properties need to be concentrated on the same block to increase a property's risk of abandonment, or do vacant properties half a mile away still have an impact? The operational definition and measurement of vacancy (whether nearby or amount in the neighborhood) have not been used consistently in existing studies of abandonment. Researchers have emphasized a variety factors, including adjacency (Myott, 1999), the same face block (Mardock, 1998; Miller, 1999) and census block group or tract (Scafidi, Schill, Wachter, & Culhane, 1998; Sternlieb, et al., 1974b). The choice of units of analysis is frequently made on the basis of data availability rather than theoretical or empirical evidence. Instead of imposing the scale of these spatial relationships by forcing them into any one areal unit such as census tracts or block groups, measures corresponding to different spatial scales can be tested to determine the range of significant spatial scales (Odland & Balzer, 1978). Raster GIS, which uses regular shaped cells to represent spatial data as a surface rather than using the awkward polygons that vector GIS depends on to represent change over space, provides a much more flexible format for measuring the concentration of nearby events and conditions. Density maps, commonly used in crime mapping, are particularly useful because they allow researchers to specify the area over which concentrations of events should be summarized, thus allowing consideration of multiple scales of spatial relationships. Scales considered in this research included adjacency (property on either side), same side of a block, face block, 100 feet, 300 feet, 500 feet, and half a mile. These area-level variables were included in the statistical model one at a time to determine scales at which these variables were statistically significant. When more than one scale was significant, the scale that generated the largest effect was included in the final model.

Predictive Value

Statistical models that predict which properties will become abandoned have practical value for the developers of early warning systems. While identifying the causes of abandonment is important for developing long-term strategies to protect housing stock, city agencies and community groups do not need to thoroughly understand the abandonment process to anticipate which properties are at greatest risk. Early warning systems should be able to indicate what combination of easily accessible indicators of abandonment have the greatest predictive value, adapting to local operational definitions of vacancy, tax arrearages, code violations, and utility terminations.

Most early warning systems have not used statistical models to determine risk levels of properties or the predictive value of indicators. Some have left this forecasting up to users, leaving them to interpret the meaning of a certain combination of data elements on their own (NKLA, Chicago NEWS). Several others have used an informal weighting system based on previous research findings and the experience of local key informants to assign properties a risk score (Mardock, 1998; Miller, 1999; Myott, 1999). A monitoring system developed by Sternlieb, et al. (1973) for Newark, NJ used analysis of variance, factor analysis, and multiple linear regression. Two other studies used discriminant analysis (Bell & Kelso, 1986) and logistic regression (Scafidi et al., 1998) to determine the magnitude and statistical significance of the relationship between indicators of abandonment and actual abandonment. The logistic regression models are more appropriate than multiple regression for analyzing dichotomous outcomes and they also generate very interpretable output in the form of odds ratios. All of these statistical models also indicate the statistical significance of relationships and report prediction error rates.

A logistic model was used in this research to analyze the predictive value of the data elements included in the NIS for a sample of properties, with a dichotomous dependent variable based on whether a property became imminently dangerous between January and November 2000. Rather than creating a sample by matching imminently dangerous properties with those that are not, properties were matched based on their location. This procedure prevented any attributes of the property from being excluded as independent variables in the predictive model. This was done by randomly selecting non-imminently dangerous properties within Philadelphia's 12 planning analysis sections in proportion to the number of imminently dangerous properties in each of them. Only residential properties were considered in this analysis, however, including houses (row, detached, and semi-detached), small apartment buildings (with no more than four units), and condominiums. Variables to be included and the order of variable entry were determined based on stepwise logistic regression using the entire data set (with 0.5 entry and 0.1 removal criterion).

Previous research offered some ideas about what variables would be most significant. Sternlieb, et al. (1973) found that properties with tax arrearages, nonwhite tenants, white owners, professional managers or rent collectors, no mortgage, adjacent deteriorated housing, and located within the urban core were more likely to become functionally abandoned, as defined by the fire and planning departments and confirmed through a foot survey. Scafidi, et al. (1998) found that properties were more likely to become financially abandoned (vested by the city of New York) if the lien-to-market value ratio exceeded 1.0, the assessed value was low, tenants took advantage of the city's emergency repair program, the property had outstanding maintenance and unsafe building violations, and the building was located in a high poverty area. Bell & Kelso (1986) found that the number of contiguous vacant parcels, the number of residential buildings and parking lots nearby, the number of units in the building, and the size of the building in relation to the parcel all helped to predict which buildings in Seattle were demolished.

Because abandonment was defined differently in these studies and because the data available to serve as predictors varied, somewhat different findings were expected for Philadelphia. All the indicators of functional, financial, and physical abandonment that are collected for the NIS were included in the initial regression models (at the individual and area level) and were expected to be significant predictors (Table 2). It was not clear which of the multiple indicators of vacancy and utility terminations would have the strongest effects and what combination of these related indicators would allow for the strongest prediction. (Bivariate correlations confirmed that many of the independent variables were highly correlated, but only two different indicators of vacancy recorded by L&I were too highly correlated to be included in the same model.)

RESULTS

Tables 3 and 4 each compare results of a single logistic regression model based on all 17,567 properties to summary results from 100 logistic regression models each based on a random sub-sample of 1000 properties. For both tables, the general results are similar in terms of the values of beta coefficients (and odds ratios). The key differences are the significance levels of these coefficients. Turning first to Table 3, notice that all variables are very significant (with P-values less than 0.01). This illustrates a general problem exhibited by regression models with large sample sizes ($n = 17{,}567$), namely their tendency to inflate significance levels. The estimates based on repeated sub-samples of size 1000 tend to reduce this effect. In addition, the dispersed nature of random sub-samples tends to reduce the effects of unobserved spatial autocorrelation, which can also inflate significance levels. The inflation of significance levels in this case result from the smoothing effect of spatial autocorrelation that tends to reduce variances of estimators, and hence to

TABLE 2

Definitions of Variables in Logistic Model of Abandonment

Indicators of Functional Abandonment	
LnIVacant	If found vacant by L&I during summer 1999 foot survey, then 1; otherwise 0
WaterSusp	If water billing was suspended (because of long-term vacancy) as of April 2000, then 1; otherwise 0
CleanSeal	If cleaned and sealed before April 1999, then 1; otherwise 0
GasShut	If gas shutoff as of January 1998, then 1; otherwise 0
POVacant	If vacant according to Post Office mail carrier as of June 1999, then 1; otherwise 0.
Indicators of Financial Abandonment	
CityOwned	If the property was owned by the city of Philadelphia in June 2000, then 1; otherwise 0
OffProp	If the owner lives off property in June 2000, then 1; otherwise 0
Mvalue	Market value (in thousands of dollars) of property as reported by the Board of Revision of Taxes in June 2000
YearsSale	Number of years since most recent sale, as of June 2000.
YearsArrears	Number of years that a property has been in tax arrearages, as of January 1, 2000
LienSale	If liens on property sold by the city during lien sale in 1997, then 1; otherwise 0
ForcedSale	If most recent sale was forced (sheriff sale, bankruptcy sale, condemnation, deed in lieu of condemnation, court ordered sales), then 1; otherwise 0
Indicators of Physical Abandonment	
HousViol	If outstanding housing code violation as of May 1999, then 1; otherwise 0
Area Indicators of Abandonment	
AdjImmDang	If adjacent property was declared imminently dangerous, then 1; otherwise 0
AdjDemo	If adjacent property has been demolished, then 1; otherwise 0
Demo_1/4	Concentration of demolished properties within a quarter mile radius
Arrears300	Concentration of tax delinquent properties within 300 foot radius
HousViol 100	Concentration of properties with outstanding code violations within 100 foot radius
OffProp300	Concentration of properties with off-property owners within 300 foot radius
CleanSeal500	Concentration of properties that have been cleaned and sealed within 500 foot radius

increase their significance levels. The mean significance levels reported for these sub-sample estimates are, therefore, probably more realistic.

It is also important to note that these 100 sub-samples yield rough estimates of the actual sampling distribution of each beta estimate in the given regression context. In principle this allows one to gauge significance levels directly without any appeal to asymptotic normality. In particular, the last column of Tables 3 and 4 reports the actual fraction of beta estimates that are less than zero. For example, the value in Table 3 of 0.04 for properties found to be vacant by the Post Office (POVacant) means that only 4 of the 100 estimates of this beta coefficient were negative. For this given regression model, then, we may conclude that there is (approximately) only a 4% chance of getting a negative

TABLE 3

Logistic Regression Output for Models with Individual-Level Data

	$n = 17{,}567$, based on 1 model			$n = 1000$, based on 100 models			
	beta coeff.	p-value	odds ratio	average beta coeff.	average p-value	average odds ratio	fraction of neg. coeff.
Lnl_Vacant	1.462	<0.001	4.314	1.555	0.002	5.043	0
WaterSusp	1.115	<0.001	3.05	1.217	0.039	3.734	0
HousViol	1.305	<0.001	3.688	1.377	0	4.125	0
YearsArrears	0.058	<0.001	1.059	0.058	0.118	1.06	0.01
CityOwned	1.352	<0.001	3.864	1.94	0.216	5.661	0.06
POVacant	0.736	<0.001	2.087	0.78	0.125	2.341	0.04
MValue	−0.028	<0.001	0.973	−0.031	0.126	0.97	0.98
OffProp	0.504	<0.001	1.656	0.443	0.225	1.64	0.08
Fire	0.602	<0.001	1.826	0.713	0.288	2.363	0.10
LienSale	0.448	<0.001	1.564	0.477	0.32	1.812	0.20
CleanSeal	0.301	0.003	1.352	0.289	0.436	1.472	0.23
YearsSale	0.007	0.003	1.007	0.008	0.375	1.008	0.25
ForcedSale	0.416	0.005	1.516	0.241	0.48	1.671	0.35
GasShut	0.213	0.008	1.237	0.111	0.475	1.189	0.38
Constant	−3.385	<0.001	0.034	−3.472	<0.001	0.034	1.00

Note. Goodness of Fit based on 1 model: −2 Log Likelikehood = 6942.737; Cox & Snell $R^2 = 0.361$; Nagelkerke $R^2 = 0.634$. Goodness of fit based on 100 models: −2 Log Likelihood = 369.684; Cox & Snell $R^2 = 0.374$; Nagelkerke $R^2 = 0.654$. Prediction results based on one model: % false positive = 2.637; % false negative = 33.283; % overall errors = 7.218. Prediction results based on 100 models: % fasle positive = 2.463; % false negative = 30.978; % overall errors = 6.646.

value for this beta estimate in a random sample of 1000 properties. It can thus be argued that although the mean P-value (0.175) indicates little significance, the sampling p-value of 0.04 provides a more meaningful indication of the positive significance of vacancy according to the post office in predicting imminent property danger. Similarly, the value 0.98 for market value (MValue) implies that there is (approximately) only a 2% chance of obtaining a positive beta estimate for this variable. Thus while the mean P-value indicates only weak negative significance of the market value, it can be argued on the basis of sampling P-values that this variable is a strong negative indicator of imminent danger for a property.

An examination of both Tables 3 and 4 shows that among the most significant variables, four stand out as the strongest determinants of imminently dangerous housing. Those properties determined to be vacant through the foot survey conducted by the L&I in 1999 (LnIVacant) are consistently at more than four times the risk of becoming imminently dangerous the following year. Properties with outstanding housing code violations (HousViol) and properties owned by the city (City Owned) have odds ratios greater than 3.5. Properties where the water department suspended billing because of long-term vacancy (WaterSusp) were also at much higher risk, with odds ratios greater than 3.0. It is also important to note in that four of the area-level variables introduced were significant, or marginally significant (Table 4). These variables included: (1) adjacent to an imminently dangerous structure, (AdjImmDang); (2) adjacent to a demolished structure (AdjImmDang); (3) number of tax delinquent properties (Arrears300); and, (4) the number of outstanding code violations (HousViol100). These results confirm that area characteristics have different scales ranging from adjacency to several hundred feet.

Prediction rates reported for all of the models are based on a classification cutoff of 0.5. False positive error rates for all the models were consistently less than 3%, indicating that

TABLE 4

Logistic Regression Output for Models with Individual and Area-Level Data

	$n = 17{,}567$, based on 1 model			$n = 1000$, based on 100 models			
	beta coeff.	p-value	odds ratio	average beta coeff.	average p-value	average odds ratio	fraction of neg. coeff.
Lnl_Vacant	1.359	<0.001	3.89	1.417	0.006	4.434	0.00
WaterSusp	1.113	<0.001	3.043	1.205	0.043	3.666	0.00
AdjImmDang	1.778	<0.001	5.917	1.963	0.001	7.917	0.00
HousViol	1.353	<0.001	3.869	1.499	<0.001	4.611	0.00
Arrears300	0.077	<0.001	1.08	0.075	0.302	1.08	0.06
YearsArrears	0.052	<0.001	1.053	0.050	0.178	1.051	0.05
POVacant	0.733	<0.001	2.08	0.784	0.145	2.398	0.05
CityOwned	1.276	<0.001	3.581	1.549	0.228	5.381	0.07
Demo_1/4	0.277	<0.001	1.319	0.264	0.283	1.335	0.13
AdjDemo	0.815	<0.001	2.259	0.846	0.221	2.641	0.04
LienSale	0.639	<0.001	1.895	0.759	0.168	2.389	0.07
OffProp	0.391	<0.001	1.479	0.365	0.32	1.538	0.16
Fire	0.614	<0.001	1.848	0.623	0.303	2.552	0.23
CleanSeal	0.452	<0.001	1.572	0.552	0.326	1.998	0.18
AdjCleanSeal	−0.427	<0.001	0.652	−0.445	0.397	0.761	0.22
ForcedSale	0.451	0.003	1.569	0.496	0.396	2.279	0.29
YearsSale	0.005	0.031	1.005	0.005	0.484	1.005	0.32
HousViol100	0.02	0.002	1.02	0.018	0.502	1.018	0.27
Mvalue	−0.008	0.004	0.992	−0.011	0.406	0.989	0.74
Constant	−4.329	<0.001	0.013	−4.434	<0.001	0.013	1.00

Note. Goodness of Fit based on 1 model: −2 Log Likelikehood = 6302.699; Cox & Snell $R^2 = 0.384$; Nagelkerke $R^2 = 0.674$. Goodness of fit based on 100 models: −2 Log Likelihood = 335.773; Cox & Snell $R^2 = 0.397$; Nagelkerke $R^2 = 0.698$. Prediction results based on one model: % false positive = 2.598; % false negative = 30.427; % overall errors = 6.759. Prediction results based on 100 models: % fasle positive = 2.312; % false negative = 28.417; % overall errors = 6.202.

properties predicted to be at high risk of becoming imminently dangerous generally did become so. However, the false negative rates are much higher (between 29 and 34%), indicating that all models fail to identify approximately one in three properties that did become imminently dangerous.

POLICY IMPLICATIONS

These results demonstrate how indicators of functional, financial, and physical abandonment can all help predict extreme physical abandonment. Similarly, different measures of the same concept can make contributions. L&I (L&IVacant), the water department (WaterSusp), and the post office (POVacant) all have different definitions of vacancy. While there is some overlap between their designations, these three measures are different enough to each contribute to the model. The results also justify, to some extent, the use of primary data collection methods. While the water department and post office record information about vacancy through routine program administration, L&I's measure of vacancy based on a foot survey contributed the most to the model.

Results also indicate that municipalities can identify properties at risk of abandonment using data that is readily available. The success of those predictions can be adjusted depending on whether the priority is to minimize false positive or false negative errors. Adjusting the classification cutoff allows one to improve the models prediction success in one direction or the other. For example, increasing the cutoff from 0.5 to 0.9 improves the false positive rate from 2.6 to 0.3% for the individual-level model. Decreasing the cutoff to 0.1 improves the false negative rate from 33.3 to 14.0%. The multivariate nature of the model makes this possible. A model including only L&I's vacancy indicator had a false positive error rate of just 5.4% and a false positive error rate of 36.8%. Without additional indicators, lowering the classification cutoff does nothing to reduce the false negative rate. These error rates have real implications for policy and resource allocation. Low false positive rates can prevent municipalities from spending scarce resources (such as funds for demolition) on properties not really at risk, while low false negative rates allow for more comprehensive, but less conservative, planning.

Using the beta coefficients from these models, it would be possible to assign a score to each property based on its likelihood of becoming imminently dangerous. Such a system could inform decisions about the appropriate intervention for different properties. Philadelphia's mayor, John Street, has proposed spending $250 million on an ambitious program to attack blight and transform neighborhoods (NTI, 2001). Under his plan, properties will be demolished, rehabilitated, encapsulated for later rehabilitation, or left alone. A risk scoring system could inform these decisions, using certain cutoff points to indicate which properties are worth investment and which are too far along in the abandonment process. Other considerations, such as proximity to schools, daycares, or other abandoned properties and their location relative to political districts, will necessarily factor into decisions about how to intervene (Research for Democracy, 2001). But multivariate statistical models like the ones presented here can provide an empirical basis for these decisions.

CONCLUSION

Early warning systems have been successful in distributing large amounts of detailed administrative housing data to wide audiences. Their developments in New York, Los Angeles, Chicago, and Philadelphia are precipitating discussions about the ethics and legality of sharing data. In addition to serving as a democratizing force, they provide an unprecedented quality control check on the information collected by city agencies, generating new feedback loops that promise to improve the quality and usefulness of municipal data. But early warning systems need to do more than just provide data. Data glut threatens to overwhelm citizens as well as the most sophisticated neighborhood-based organizations and city agencies. The neighborhood indicators movement (Sawicki & Flynn, 1996) has challenged researchers to help community groups and residents interpret the meaning of raw data so that it is usefully for planning and evaluation. By organizing large amounts of related information through simple reports and maps, early warning systems provide a value-added product. But they need to move beyond offering descriptions and visualizations of property and neighborhood conditions into the realm of spatial and temporal statistical data analysis.

While logistic regression provides a clear improvement over more ad hoc approaches (such as weight systems based on key informants that do not quantify their error rates), it is limited in its ability to predict housing abandonment because it forces relationships into a linear model. Possible alternatives include hazard models that could analyze the time to failure (abandonment) of properties or flexible-functional-form models such as neural networks that could detect a broader range of relationships among variables.

The methods employed in this research are transferable to other cities, but regardless of the type of statistical model, the outcome and predictor variables will need to reflect local data availability and concerns. The most important decision facing Philadelphia is whether or not to demolish a property, and knowing whether it will become imminently dangerous or not can facilitate this decision. In other cities that have clear policies for *in rem* housing (such as New York, Atlanta, and Cleveland), the point at which a property can legally be declared financially abandoned and acquired by the city may be a more appropriate focus. Once predictive models are developed, they need to be tested and refined as conditions and definitions change over time. While these types of early warning systems offer no panacea to the difficult and expensive problems caused by large-scale housing abandonment, they do provide support for making sound and consistent decisions about the fate of distressed properties.

REFERENCES

Accordino, J., & Johnson, G. T. (2000). Addressing the vacant and abandoned property problem. *Journal of Urban Affairs*, *22*(3), 301–315.

Arsen, D. (1992). Property tax assessment rates and residential abandonment: Policy for New York City. *American Journal of Economic Sociology*, *51*(3), 361–377.

Bailey, T. C., & Gatrell, A. C. (1995). *Interactive spatial data analysis*. Essex, England: Longman Scientific & Technical.

Bell, E. J., & Kelso, D. (1986). The demolition of downtown low-income residential building: A discriminant analysis. *Socio-Economic Planning Science*, *20*(1), 17–23.

Bender, B. (1979). The determinants of housing demolition and abandonment. *Southern Economic Journal*, *46*(2), 131–144.

Brown, R. L. (1999). Vacant Housing in the Context of a City Neighborhood: The West Side, Saint Paul, Minnesota. Saint Paul: Neighborhood Planning for Community Revitalization, Center for Urban and Regional Affairs, University of Minnesota.

Costa, F. J., & Hanten, E. W. (1988). Creating an Early Detection System for Housing Abandonment in Akron, Ohio. Akron: Center for Urban Studies, The University of Akron.

Cressie, N. C. (1993). *Statistics for spatial data*. New York: John Wiley & Sons, Inc.

Galster, G. (1987). *Homeowners and neighborhood reinvestment*. Durham: Duke University Press.

Grigsby, W., Bratz, M., Galster, G., & Maclennan, D. (1987). The dynamics of neighborhood change and decline. In D. Diamond & J. B. McLoughlin (Eds.), *Progress in Planning* (Vol. 28, Part I) (pp. 7–73). Pergamon Press.

Keenan, P., Lowe, S., & Spencer, S. (1999). Housing abandonment in the inner cities: The politics of low demand for housing. *Housing Studies*, *14*(5), 703–716.

Kiefer, D. (1980). Housing deterioration, housing codes and rent control. *Urban Studies*, *17*, 53–62.

Mardock, L. (1998). *Predicting housing abandonment in Minneapolis' central neighborhood: Creating an early warning system*. Available: http://www.npcr.org/reports/npcr1089/npcr1089.html [September 2002]. Minneapolis, MN: Neighborhood Planning for Community Revitalization.

Miller, J. (1999). *Developing an early warning system for an eight-block area of Minneapolis' Phillips neighborhood: Methodology, results, and implications for future studies*. Available: www.npcr.org/reports/npcr1138/npcr1138.html [September 2002]. Minneapolis, MN: Neighborhood Planning for Community Revitalization.

Myott, Eric. (1999). *Early warning system feasibility in the Hamline Midway area*. Available: www.npcr.org/reports/npcr1121/npcr1121.html [September 2002]. Minneapolis, MN: Neighborhood Planning for Community Revitalization.

Neighborhood Transformation Initiative. (2001). *A strategy for investment and growth*. Available: www.phila.gov/mayor/jfs/mayorsnti/strategy/nti_final_presentation2.pdf [September 2002]. Neighborhood Transformation Initiative, Philadelphia, PA.

O'Flaherty, B. (1993). *Abandoned Buildings: A Stochastic Analysis*. *Journal of Urban Economics*, *34*, 43–74.

Odland, J., & Balzer, B. (1978). Localized externalities, contagious processes and the deterioration of urban housing: An empirical analysis. *Socio-Economic Planning Sciences*, *13*, 87–93.

Odland, J., & Barff, R. (1982). A statistical model for the development of patterns: Applications of the spread of housing deterioration. *Geographical Analysis*, *14*(4), 326–339.

Philippi, B. (1977). Neighborhood effects and maintenance of urban housing stock. *Journal of Mathematical Sociology*, *5*(2), 151–172.

Research for Democracy. (2001). *Blight free Philadelphia*. Philadelphia: Temple University.

Sawicki, D. S., & Flynn, P. (1996, Spring). Neighborhood indicators: A review of the literature and assessment of conceptual and methodological issues. *Journal of the American Planning Association*, *62*(2), 165–183.

Scafidi, B. P., Schill, M. H., Wachter, S. M., & Culhane, D. P. (1998). An economic analysis of housing abandonment. *Journal of Housing Economics*, *7*, 287–303.

Sternlieb, G., & Burchell, R. W. (1973). *Residential abandonment: The tenement landlord revisited*. New Brunswick, NJ: Center for Urban Policy Research.

Sternlieb, G., Hughes, J. W., Bleakly, K., & Listokin, D. (1974a). *Housing abandonment in Pennsylvania*. New Brunswick, NJ: Center for Urban Policy Research.

Sternlieb, G., Burchell, R. W., Hughes, J. W., & James, F. J. (1974b). Housing abandonment in the urban core. *Journal of the American Institute of Planners*, *40*(5), 321–332.

Villares, G. (1999). Strategies in Dealing with Problem Properties in Dayton's Bluff. Saint Paul: Neighborhood Planning for Community Revitalization, Center for Urban and Regional Affairs, University of Minnesota.

Wilson, D., Margulis, H., & Ketchum, J. (1994). Spatial aspects of housing abandonment in the 1990s: The Cleveland experience. *Housing Studies*, *9*(4), 493–510.

White, M. J. (1986). Property taxes and urban housing abandonment. *Journal of Urban Economics*, *20*, 312–330.