Geographical distribution of patients visiting a health information exchange in New York City

Arit Onyile,1,2 Sandip R Vaidya,2 Gilad Kuperman,2,3 Jason S Shapiro1,2

ABSTRACT
Objective For a health information exchange (HIE) organization to succeed in any given region, it is important to understand the optimal catchment area for the patient population it is serving. The objective of this analysis was to understand the geographical distribution of the patients being served by one HIE organization in New York City (NYC).

Materials and Methods Patient demographic data were obtained from the New York Clinical Information Exchange (NYCLIX), a regional health information organization (RHIO) representing most of the major medical centers in the borough of Manhattan in NYC. Patients’ home address zip codes were used to create a research dataset with aggregate counts of patients by US county and international standards organization country. Times Square was designated as the geographical center point of the RHIO for distance calculations.

Results Most patients (87.7%) live within a 30 mile radius from Times Square and there was a precipitous drop off of patients visiting RHIO-affiliated facilities at distances greater than 100 miles. 43.6% of patients visiting NYCLIX facilities were from the other NYC boroughs rather than from Manhattan itself (31.9%).

Discussion Most patients who seek care at members of NYCLIX live within a well-defined area and a clear decrease in patients visiting NYCLIX sites with distance was identified. Understanding the geographical distribution of patients visiting the large medical centers in the RHIO can inform the RHIO’s planning as it looks to add new participant organizations in the surrounding geographical area.

BACKGROUND AND SIGNIFICANCE
Through the Health Information Technology for Economic and Clinical Health (HITECH) Act, which is part of the American Recovery and Reinvestment Act of 2009, the US government has taken an unprecedented step towards promoting interoperability through secure electronic health information systems nationwide, including the allocation of US$19.2 billion towards the adoption and meaningful use of health information technology (HIT) and US$564 million for state-level health information exchanges (HIE).

The goal of HITECH is to improve the quality of health for Americans through the adoption and meaningful use of HIT. In order for this to be accomplished successfully, and for the full potential of HIT to be realized, an increased emphasis is placed on interoperability between healthcare providers and provider organizations, as evidenced in the meaningful use stage 2 proposed rule. With along with directed exchange initiatives and consumer-based HIE efforts, state-level HIE funded by HITECH will help improve interoperability and move towards a nationwide HIE network.

Before the federally funded state-level HIE programme, and often serving as its building blocks, regional health information organizations (RHIO) bring together multiple stakeholders within a given geographical area, managing governance, policy development and operational issues as they create HIE at a regional level. Although there has been some regular study of the 255 functioning HIT currently in existence nationwide, to date there is a paucity of analysis in the literature describing the geographical distribution of patients within a RHIO in a metropolitan area. We hypothesize that a RHIO in a major metropolitan area would have a concentration of patients visiting its participant organizations from within the city, and then a ‘penumbra’ effect with patients visiting from outlying boroughs and suburbs, with a likely inflection point in visits per capita as distance from the center of the RHIO increased. The area in this ‘penumbra’ or ‘white space’ may contain provider organizations and other potential stakeholders that are not part of the RHIO. For one RHIO in New York City (NYC), this study aims to identify the ‘catchment’ area of the RHIO by identifying how the proportion of patients visiting the RHIO varies with distance and other geographical factors. Identifying this catchment area and understanding the geographical distribution of patients will aid in the strategic planning process when seeking to align with new partner organizations in the region.

The objectives of this study were: (1) to identify the proportion of patients in the HIE within the primary service area of the participant organizations of the RHIO; (2) to identify the proportion of patients from the immediately contiguous geographical areas, nationwide and internationally; and (3) to determine if there was an inflection point that indicated the primary area of the coverage of the RHIO, thereby defining the ideal catchment area for the RHIO.

MATERIALS AND METHODS
Setting At the time of the study (2011), there were four functioning RHIO in NYC, one in each of the four largest boroughs: Manhattan, Queens, Brooklyn and The Bronx. A fifth RHIO covered Long Island, a region adjacent to NYC, and a sixth RHIO covered Westchester County, also adjacent to NYC.

The New York Clinical Information Exchange (NYCLIX) was a Manhattan-based RHIO, which built an operational, secure HIE available to authorized hospitals and other healthcare providers.
in the NYC metropolitan area. (At the time of this study, NYCLIX was an independent HIE. As of April 2012, NYCLIX merged with the Long-Island-based RHIO Lipix to form a new organization, Healthix Inc.) NYCLIX was a collaboration among ambulatory physician groups, long-term care facilities, a Medicaid managed care plan, the nation’s largest home health-care provider and academic medical centers that serve as major referral centers with a total of 7503 inpatient beds, 341,065 annual inpatient discharge and 540,854 annual emergency department visits (see table 1). NYCLIX began operations in 2007 and its stakeholders were in Manhattan, Brooklyn and Staten Island.

Datasets
In order to understand the geographical distribution of patients visiting a NYCLIX facility, we used data from the NYCLIX master patient index (MPI). The MPI is a database that maintains an identifier for every unique patient registered within the HIE. The MPI receives demographic information from each participant organization and includes patient address and zip code and/or country name. In the analysis, when the patient had address information from multiple sources, the most recent source was used. NYCLIX began to operate functionally in March 2009 and contains information on patients who visited a NYCLIX participant since that time. The NYCLIX operations team had access to full US zip codes and mapped the most current zip code for each unique patient to the appropriate US county using an internal SAS dataset, SASHELP.zipcode. Times Square was designated as the NYCLIX geographical center point. The operations team then calculated the distance from each zip code to Times Square. Any patient who did not have a five or nine-digit zip code was excluded from the study. The final dataset that was delivered to the research team included aggregate counts of patients by US county and international standard organization (ISO) country code and distance from Times Square. This dataset was imported into SAS V9.2 for further analysis.

The protocol for this study was reviewed by the Mount Sinai Program for the Protection of Human Subjects and was issued a determination letter stating it is not human subjects research.

Statistical methods
The number of unique patients visiting a NYCLIX facility was the unit of measure. US Census Bureau data were used for denominators in calculating adjusted rates for specific geographical regions (number of patients from county divided by the number of people living in the county). For each county, the rate of people visiting a NYCLIX facility was calculated by dividing the number of people who had visited a NYCLIX site by the total number of people in the county (2010 census counts). Charts, graphs and maps using aggregate counts and rates were created for counties, states and geographical regions within the USA and ISO countries (see figures 1–3 and tables 1–4). We identified the ISO countries in the research dataset and countries that could not be mapped were manually coded by the research team whenever possible.

EpilInfo V3.5.3 was used to create state and regional maps. Data from the US Census Bureau was used in running spatial regression analyses. Spatial regressions were performed using SatScan V9.1.1 released 9 March 2011 to identify spatial clusters at the state and county level and determine the likelihood or relative risks (RR) that residents in these spatial clusters would visit a NYCLIX facility by comparing the NYCLIX data to census data. A cluster county was defined as a county that had a RR greater than 1 (which is considered statistically significant). Residents from those counties are more likely to visit a NYCLIX facility when compared to residents from other counties. The following surrounding tri-state counties were excluded for the spatial regression: New York: NYC, Nassau, Suffolk, Westchester; New Jersey: Bergen, Essex, Hudson, Middlesex, Monmouth; and Connecticut: Fairfield and New Haven. Doing this allowed us to look for cluster counties not in the immediate NYCLIX geographical area. The spatial regressions helped examine the relationship between the likelihood of visiting a NYCLIX facility (dependent variable) and geographical/spatial location (independent variable).

RESULTS
As of 27 May 2011 there were 4,006,605 unique patients who had visited a NYCLIX facility or provider. Of these, 0.7% (26,589) patients were excluded from the analysis for having an invalid (non-five or nine-digit) zip code leaving 3,980,016 for analysis. Our analysis plan consisted of analyses grouped into three geographical categories: immediate NYCLIX geographical area, nationwide and internationally.

Tri-state (New York, New Jersey, Connecticut)
New York, New Jersey and Connecticut represented 3,867,738 patients out of a total of 3,980,016 in the NYCLIX dataset (97.2% of the 3,980,016 of those in NYCLIX; with a visit rate of 12.1 per 100 patients living in those states based on 2010 census data) (table 2). Table 3 depicts the incremental increase in the number of patients comprising the total patients in NYCLIX with distance. The rate of patients visiting NYCLIX per capita dropped precipitously with distance from 24.3 per 100 within 30 miles to 0.42 per 100 at 100 miles (figure 1); 3,863,854 patients were in counties within 100 miles from the NYCLIX geographical center, Times Square, representing 97.1% of the total patients in NYCLIX. Within the range of 0–100 miles, the average rate of patients visiting a NYCLIX facility was 8.7 per 100; 87.7% of patients in NYCLIX (3,487,499) live within a 30 mile radius from Times Square and the inflection point, after which the rate of patients visiting NYCLIX remains below one per 100, was at approximately 80 miles from Times Square (see figure 1).

| Table 1 | Data related to NYCLIX member hospitals during the study period |
|---------|---------------------------------------------------------------|
|          | Beds\(^13\) | Inpatient discharges\(^14\) |
| **Sites in Manhattan** |                  |                              |
| Site 1   | 899        | 41186                         |
| Site 2   | 1171       | 59063                         |
| Site 3   | 814        | 45404                         |
| Site 4   | 1178       | 50675                         |
| Site 5   | 879        | 31546                         |
| Site 6   | 505        | 29333                         |
| Site 7   | 523        | 19903                         |
| **Sites in other boroughs of NYC** |                  |                              |
| Site 8   | 714        | 41819                         |
| Site 9   | 376        | 19057                         |
| Site 10  | 212        | 11258                         |
| **Totals** | 7503       | 341065                        |

NYC, New York City; NYCLIX, New York Clinical Information Exchange.
New York City
A total of 3,004,699 patients resided in NYC (Bronx, Kings (Brooklyn), Manhattan, Queens and Richmond (Staten Island) counties), representing 75.5% of patients in NYCLIX and 35.6 rate per 100 people in this geographical area (table 2). When combined, the absolute count of patients visiting NYCLIX sites from the four NYC outer boroughs excluding Manhattan was 1,735,783 (43.6% of NYCLIX patients). This is larger than the number of patients visiting NYCLIX sites from Manhattan itself, 1,268,916 (31.9% of NYCLIX patients) (figure 2), although the rates of patients visiting a NYCLIX site per capita were lower in the other boroughs (25.5 rate per 100) when compared to Manhattan (77.8 rate per 100). Within NYC, when excluding Manhattan, Brooklyn had the highest absolute count of patients, 718,727 (18.1% of NYCLIX patients; 27.8 rate per 100). When excluding NYC and taking into account 2010 census population counts, the areas surrounding NYC (Long Island: Nassau, Suffolk; and Westchester counties) had the highest rate of patients visiting NYCLIX facilities, 9.5 rate per 100 (figure 3). A total of 364,023 patients (9.1% of NYCLIX patients; 9.5 rate per 100) visited a NYCLIX facility from these regions. The regions surrounding the NYC metropolitan area were identified as spatial clusters. The residents in these regions had a greater likelihood of visiting a NYCLIX facility than not visiting a NYCLIX facility when excluding NYC.

The New York downstate region, made up of NYC, Long Island, Hudson Valley and Catskills regions of New York state had the highest rate of patients visiting a NYCLIX facility (3,526,611; 88.6% of NYCLIX patients; 24.8 rate per 100).

Surrounding tri-state counties (New York, New Jersey, Connecticut)
The surrounding tri-state NYCLIX counties were identified as being New York: Nassau, Suffolk, Westchester; New Jersey: Bergen, Essex, Hudson, Middlesex, Monmouth; and Connecticut: Fairfield and New Haven. These counties accounted for 600,312 patients in our dataset who visited a NYCLIX provider (15.1% of NYCLIX patients; 6.3 rate per 100) compared to 262,727 patients from the remaining counties within the tri-state region (New York, New Jersey, Connecticut) (6.6% of NYCLIX patients; 1.8 rate per 100).

Cluster analysis
When excluding NYC and the nearby surrounding counties the following counties were identified as cluster counties in New York: Duchess, Orange, Putnam, Rockland, Sullivan, Ulster; and New Jersey: Morris, Somerset, Sussex and Union. The RR for patients living within these counties to visit a NYCLIX site compared to the rest of the state is 14.40 (p<0.05) when excluding the following counties: New York: NYC, Nassau, Suffolk, Westchester; New Jersey: Bergen, Essex, Hudson, Middlesex and Monmouth; and Connecticut: Fairfield and New Haven (see table 2).

Nationwide analysis
All 50 states, District of Columbia and United States territories had representation in our dataset. As of April 2011, as per the US Census Bureau there were 41,357 US zip codes and 1931 US counties. The NYCLIX patient population had representation in 77.7% (1500) of all US counties and 39.7% (16,401) of all US counties. *Includes Manhattan, Brooklyn, Bronx and Queens and Staten Island.
‡Includes Nassau, New York and Suffolk, New York counties.
†Excludes NYC, Westchester, Nassau and Suffolk counties.
NYC, New York City; NYCLIX, New York Clinical Information Exchange.
zip codes and 890 (97.4%) of the 914 three-digit zip code prefixes that currently exist in the USA.

International analysis
As per the MARNAMES SAS dataset, there are 239 unique ISO countries worldwide and the NYCLIX patient population spans across 57 ISO countries (excluding the USA); 1739 patients (0.04% of NYCLIX patients) came from overseas (see table 4).

DISCUSSION
NYCLIX has representation in all 50 US states, four US territories and 57 ISO countries. With the diversity of the New York metropolitan area it is not surprising that a RHIO situated in this region has such a wide representation of patients.

As shown in figure 3 there are a set of concentric circles around NYC depicting the decreasing numbers of patients visiting NYCLIX with distance. Most patients come from the regions surrounding NYC. Although our analysis does not show anything we did not expect, it confirms our initial hypothesis. There is a relationship between a patient’s permanent address and distance traveled to see a healthcare provider. Also as predicted, the concentration of patients visiting the RHIO sites per capita decreased with distance, but not in a linear fashion, with an inflection point in the penumbra of the RHIO, seen in figure 1 between 60 and 100 miles from the center of the RHIO. This information has aided the RHIO in strategic planning when deciding which stakeholder organizations to approach when it comes to expansion, and similar approaches may be helpful to other RHIO across the country.

One surprising result was that the outer boroughs combined, Bronx, Brooklyn (Kings County), Staten Island (Richmond) and Queens, have a larger absolute count of patients visiting NYCLIX than Manhattan, where most of the RHIO provider organizations are based. This suggests that the behavior of patients seeking healthcare is fluid. Patients move throughout the local region for healthcare. Currently, there are six RHIO covering NYC, Long Island and Westchester, which appear to have taken the problem of fragmentation of patient information that existed at the provider organization level, and aggregated it up to the level of the RHIO. While aggregation of data at this higher level of the overall healthcare system is an improvement, it still amounts to silos of data given the level of patient fluidity shown here. This suggests a need for inter-RHIO interoperability perhaps using the nationwide health information network’s query-based exchange model that relies on aggregate data being accessible and in turn provides data access for unplanned care and when a patient travels across provider organizations. Without this level of interoperability, it will be difficult or impossible for any of the current individual RHIO to reach a critical mass of patients or data in order to provide comprehensive higher level services such as care coordination or quality measurement.

The NYC metropolitan area has a mono-centric city pattern of commuting, with many workers commuting to a central point, Manhattan, but residing in suburban areas. Most patients reside within 60 miles from Times Square. Although there is an overall negative relationship between distance and the likelihood of visiting a NYCLIX provider, there are some peaks and valleys between 30–100 miles from Times Square. At
distances greater than 100 miles from Times Square the rate of visits per capita is very small. This may indicate that some NYCLIX patients include commuters into Manhattan who choose to seek care close to their place of work rather than where they reside, or may suggest that even though some of the participants in NYCLIX are major referral centers, that the fraction of patients that constitute referrals from a distance is not that significant.

Cluster counties are identified as regions where residents have a higher likelihood of visiting a NYCLIX site. We ran a Poisson regression to identify cluster counties. As we excluded counties close to NYC the counties in the next most distal region became cluster counties. We did not find any cluster counties as outliers at large distances from NYC, which we had considered a possibility. Our findings suggest that the likelihood of visiting a NYCLIX provider is primarily dependent on distance. Furthermore, even within a dense metropolitan area with large academic medical centers and specialists, patients are likely to seek healthcare as a function of distance from where they reside or work.

Limitations
In calculating the rates, the numerator and denominator came from different datasets. The numerator was a dynamic number representing the cumulative total of patients who have visited a NYCLIX site or provider from March 2009 to May 2011, and some patients may not have been living at the listed home address at the time the census data were collected. The denominator was a static number representing the census count at a specific point in time in 2010. There is thus a possibility that the rates calculated do not reflect the true rate as of 2011. Fifteen out of the 57 country entries found in the NYCLIX dataset did not match to an ISO country name in the MAP NAMES SAS dataset. These countries had to be manually mapped by the research team. Finally, the MAP NAMES SAS dataset, used to identify ISO countries, was released in December 2008 and may not accurately reflect the current number of ISO countries as of the date of the analysis.

CONCLUSIONS
This study shows that patients are highly mobile and visit providers across city, county, state and national lines. Under HITECH there will be a transition from paper medical records to electronic medical records. RHIO and HIE offer the opportunity for healthcare data to be shared among providers and facilities across city, county and state jurisdictions, more accurately reflecting the way in which patients actually access healthcare. Although NYC represents the largest metropolitan area in the country, and it is not clear that these results will generalize to other regions, similar analyses may aid other regions in determining how best to spread scarce resources when deciding where the geographical boundaries of an HIE organization should be drawn.

Acknowledgements The authors extend special thanks to the NYCLIX operational team, Thomas Moore and Laurence Berg, for their role as honest brokers in running the database queries for this study.

Contributors The four authors are justifiably credited with authorship according to the authorship criteria. AO: conception, design, analysis and interpretation of data, drafting of the manuscript, final approval of the version to be published; SRV:...
conception and design, acquisition of data, drafting of the manuscript, final approval of the version to be published; GK and JSS: conception and design, interpretation of data, drafting of the manuscript, final approval of the version to be published.

**Funding**  This work was supported by the National Library of Medicine grant number R00LM009556.

**Competing interests**  None.

**Provenance and peer review**  Not commissioned; externally peer reviewed.

**Open Access**  This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

**REFERENCES**

1. Blumenthal D. Launching HITECH. *N Engl J Med* 2010;362:382–5.
2. Steinbrook R. Health Care and the American Recovery and Reinvestment Act. *N Engl J Med* 2009;360:4.
3. Nationwide Health Information Network (NHIN). Background & Scope. Washington, DC, 2009: http://www.hhs.gov/healthit/healthnetwork/background/ (accessed 14 May 2012).
4. 2011 Report on Health Information Exchange: The Changing Landscape. eHealth Initiative. 2011 December 2010.

5. Report to the President Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward. 2010 December 2010.
6. Medicare and Medicaid Programs: Electronic Health Record Incentive Program-Stage 2. In: Services DoHHSCHM, editor. 2012. p 132.
7. Thomewill IDAF, Cox BA, Estethay RJ. Information infrastructure for consumer health. *Am J Prev Med* 2011;40:5123–533.
8. Kuperman GJ. Health-information exchange: why are we doing it, and what are we doing? *J Am Med Inform Assoc* 2011;18:678–82.
9. Williams C. Update on the State HIE Program. 2012. healthit.hhs.gov. (accessed 5 Sept 2011).
10. New National Survey Shows Increased Privacy Controls. Concerns with systems integration and participation in accountable care. Washington, DC: eHealth Initiative, 2011.
11. Williams CMF, Mertz K, Hegin E, et al. From the Office of the National Coordinator: the strategy for advancing the exchange of health information health affairs. 2012;31:527–36.
12. Bram JMA. The evolution of commuting patterns in the New York city metro area. *Curr Issues Econ Finance* 2005;11:6.
13. New York State Hospital Profile. New York State Department of Health. 2009 June 29: http://hospitals.nyhealth.gov/index.php?PHPSESSID=51f042212df5db2e0ff1540164779 (accessed 9 Sep 2009).
14. Hospital Inpatient Data of New York State. 2009 Annual Report from the Statewide Planning and Research Cooperative System (SPARCS). New York State Department of Health 2010: http://www.health.ny.gov/statistics/sparcs/annual/ip/2009/ t2009_08.htm (accessed 16 Jul 2012).