Evaluation of the Integrated Care Programme in Waltham Forest, East London and City (WELC), England

18 April 2016

Arne T. Wolters and Adam Steventon

The Health Foundation

EXECUTIVE SUMMARY

This protocol outlines a proposed quantitative evaluation of the Waltham Forest, East London and City (WELC) Integrated Care Programme. The aim of the study is to test the feasibility of using matched control within rapid cycle evaluation. In the short term, this work is constrained by the data that are currently available, and hence we are focussed on identifying changes in emergency hospital admissions, the number of hospital bed days per patient, and the nominal cost of primary and secondary care per patient. Should this approach be successful, it could be extended to incorporate new data collections on patient-reported outcomes and the experience of care.

The study would focus on patients enrolled during 2014 who were in the top 20% in terms of their risk of emergency admission and aged fifty years or over. Although the approach could be extended to look at other patient groups, we are restricting the initial work to this group since it comprises 85% of all patients enrolled onto the integrated care pathway in 2014 and, from an information governance perspective, we would like to minimise the amount of data that needs to be transferred to the evaluation team. We will compare the endpoints of these patients with those of retrospectively matched control patients from within the WELC area.

The evaluation will begin in September 2015, subject to appropriately pseudonymised, line-level patient information being transferred to the secure, accredited data processing environment at the Health Foundation in London. We have been collaborating closely with the WELC steering group to ensure that this project delivers results that are timely with a view to operational decisions about the programme. Provisional results, based on a six-month follow-up period, are expected by January 2016. Following a second wave of line-level patient information, a full report based on twelve months of follow-up is expected to be ready in April 2016. Findings will then be submitted for publication in a peer-reviewed journal.

To evaluate the integrated care programme the Health Foundation requires access to pseudonymised patient-level information including all patients in the top twenty per cent at risk of emergency hospital admission aged over fifty in the WELC area. Read-coded GP data on registration with GPs, prescriptions, diagnoses, and encounters, together with line-level data from the admitted patient care, outpatient and accident and emergency commissioning data set are required to evaluate the programme. All pseudonymised data will be held on the Health Foundation's secure processing environment for the duration of the project.

This protocol was approved by the WELC Evaluation Steering Group on 16 June 2015.

Update of document: The protocol has been updated on 11 January 2015 to reflect new timelines, and the exclusion of the Waltham Forest area from the feasibility study and preliminary results.

Update of document: The protocol has been updated on 18 April 2016 to reflect the new timelines after the delay in data access permissions and file transfer.

CONTENTS

INTRODUCTION

The Waltham Forest, East London and City (WELC) Integrated Care Programme is focused on revolutionising care for a population of almost one million people in an area facing significant health and social challenges. The Boroughs of Newham, Tower Hamlets and Waltham Forest are some of the most deprived in London, with significant health needs and inequalities. The programme is made of up Clinical Commissioning Groups (CCGs) and Councils in the three Boroughs along with Barts Health NHS Trust, North East London Foundation Trust, East London Foundation Trust and UCL Partners. It aims to facilitate a joint health, social care and mental health approach.

Patients at high risk of emergency hospital admission (top quintile) or with long-term conditions are offered enrolment in the programme during the course of a consultation with their General Practitioner (GP). Enrolment on the programme is considered complete if patients consent to be enrolled on the programme and consent for their data to be shared with the WELC Integrated Care Programme partners, and this consent is recorded on the electronic patient record by the GP. Enrolment started in April 2012, with a total of 29 patients enrolling in the first 12 months. The following fiscal year (2013-2014) saw another 906 patients enrol on to the programme. Numbers are increasing rapidly, with currently 11,940 patients on the pathway (1.3% of all patients, in comparison to the 20% targeted) (source: WELC Data Cube).

The integrated care programme is trying to achieve its goals through a combination of patient engagement; joint decision making and accountability; clinical leadership and culture development; information sharing and decision support; and aligned incentives and reimbursement models. Although patients are either enrolled or not enrolled on the programme, the actual intervention may differ from patient to patient and across different CCGs. The Integrated Care Programme comprises nine categories of interventions:

1. Self-care, behaviour and expectation management;
2. Care planning;
3. Health and social care navigation;
4. Case management;
5. Specialist input in the community;
6. Discharge support for mental health patients from secondary to primary care;
7. Rapid response with short term re-ablement;
8. Mental health liaison; and
9. Discharge support from the acute community.

The exact interventions can be specific to local arrangements, and so are difficult to generalise across a larger population. Therefore, the evaluation proposed in this protocol aims to evaluate the overall effect of enrolment onto the integrated care pathway, rather than looking at effect of the individual

interventions offered during this pathway. We will, however, try to describe the occurrence of particular interventions.

We propose to evaluate the integrated care programme using a (matched) control group, rather than solely tracking metrics using the data cube for intervention patients. Looking at patient level data, and matching patients enrolled in the programme with similar patients who were not enrolled, allows us to compare the outcome measures of enrolled patients with a counterfactual. This is particularly important given the tendency of patients with a history of hospital admissions to see reductions in admissions over time even in the absence of a specific intervention ('regression to the mean') (Roland et al. 2005).

The first step of the evaluation would be to assess the feasibility of the matched control approach. There are risks that we might not find controls that are adequately matched to enrolled patients on observed baseline variables, or that there might be unobserved confounding (*e.g.*, as indicated by differential mortality rates between enrolled patients and the control group). Provided adequate controls can be found, we propose to evaluate the programme as described below.

METHODS

Study cohort

We will study patients enrolled onto the integrated care pathway in 2014 (*i.e.*, between 1 January 2014 and 31 December 2014) who were in the top 20% in terms of their risk of emergency hospital admission and who were aged fifty years or over. The cohort will include patients from all three CCGs, though we will also analyse the patients in the three CCGs separately.

This cohort has been chosen to ensure a sufficiently large study cohort of at least 6,360 patients (*i.e.*, the difference between the total number of patients enrolled on 1 January 2015 and 1 January 2014, source: WELC Data Cube). The total number of patients in the cohort is expected to be higher than this, as enrolled patients who died, or moved out of the three CCGs are not included in this total.

This cohort includes 85 per cent of patients currently enrolled in the programme. By limiting the study to this cohort ensures that sufficient information is used to evaluate the effect of the programme for different age groups robustly. Evaluation of the programme for younger patients, or patients that are less at risk of emergency hospital admission would add too much uncertainty to the study's findings at this stage. Limiting the cohort also allows for the matching to be competed in a timely manner, as matching algorithms used are very time consuming.

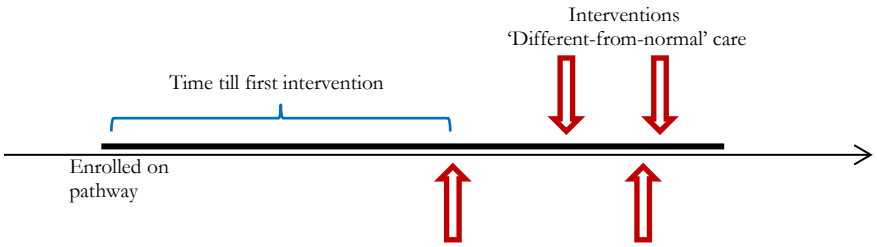
A small number of patients provided consent to be on the pathway or for their data to be shared but not both. Because there is some ambiguity about the care received by these patients, they will be excluded from the study.

The initial findings of the evaluation will be based on outcomes over the six months following enrolment. This relatively short follow-up period will allow for timely results by the end of 2015. Further analysis will be based on a follow-up period of twelve months following enrolment; however patient-level data required for this analysis are not expected to be available before February 2015 (see 'Timescale' below for more details).

Intervention

The focus of this study is to evaluate the overall integrated care pathway, and thus we will assess patient outcomes from the date of enrolment onto this pathway. After the enrolment on the programme a patient will receive specific interventions at varying intervals that greatly influence the effect of the overall 'treatment' (see figure below).

We will report the overall duration of time spent on the integrated care pathway using Kaplan-Meier curves (Kaplan & Meier 1958). These show the proportion of patients remaining on the pathway after various numbers of days spent on the pathway. In some cases, the lag between enrolment on the pathway and receiving first 'different-from-normal' care might be substantial, and so we will explore whether the individual interventions illustrated in the figure below can be identified from the line-level data.



Endpoints

The Integrated Care Program has identified a wide range of Key Performance Indicators (KPIs). These indicators span: quantitative metrics to assess how care is being integrated; metrics on reviews of individual patient cases and care plans; and metrics of health care utilisation. This last category of indicators is most relevant to this study and includes:

- Number of GP appointments (*e.g.*, per head, per 12 months);
- Number of urgent referrals;
- Number of hospital bed days per patient;
- Number of weekend discharges from hospital;
- Secondary care cost (PbR);
- Number of emergency hospital admissions;
- Number of avoidable hospital admissions;
- Number of A&E attendances; and
- 30-day hospital readmission rates.

Our primary endpoint is the number of emergency hospital admissions per patient, which is a commonly used endpoint in evaluative studies of integrated care. Our secondary endpoint is the number of hospital bed days per patient over twelve months, as this reflects changes to the length of stay in hospital as well as the number of admissions.

Other endpoints will be the notional cost of care per patient, including primary and secondary care cost and other indicators listed above.

Sample size calculation

Our sample size will be sufficient to identify a 15% change in emergency admissions between the enrolled patients and the matched control group. Our sample size calculation was performed in SAS with 90% power and a two-sided p-value of 0.05. We assumed an average number of emergency admissions per patient per year of 0.6 with a standard deviation of 1.4. These assumptions are based on a previous study with a similar population base (Steventon, Bardsley, Billings, Dixon, et al. 2012).

The sample size calculation indicate that in order to identify a 15% change in emergency admission a sample size of 10,174 is required in case of a balanced sample (1 control per enrolled patient), or 11,445 in case of a sample with two controls per enrolled patient. In comparison, we estimate that our study will include at least 6,360 integrated care patients. In a matched control design with one control patient per integrated care patient, the total sample size would therefore be 12,720 patients. If there were two control patients per integrated care patient, the total sample size would be higher, at 19,080.

Control group

In the absence of a randomised controlled trial, we will compare the outcomes of the integrated care cohort with a retrospectively matched control group. In other words, line-level routine, patient information will be used to characterise the integrated care patients as well as possible, and then select, for every patient, matched controls that were not enrolled onto the programme but had similar observed characteristics to the enrolled patient. These matched patients will form the control group for the evaluation of the programme.

Control patients will be registered with a general practice in the WELC area, and will not have been enrolled onto the integrated care pathway at any point before the end of our follow-up period. Preliminary analysis of aggregated data from the WELC Integrated Care Programme (using the Data Cube) suggests that the pool of potential matches is sufficiently large to select matched controls within cluster (*i.e.*, geographically clustered GPs) and within the same risk-band (top 0.5%, 0.5-5%, or 5-20%). This is because enrolment rates at cluster level vary between 17% and 38%, and enrolment rates for the different risk-bands are 50%, 24% and 2%, respectively.

Matching will be done to optimise the similarity of integrated care and matched control groups with respect to variables that are likely to be predictive of the endpoints. These variables will include:

predictive risk score; age; gender; ethnicity; prior primary and secondary care utilisation; a measure of socioeconomic deprivation (the Index of Multiple Deprivation, attributed to the Lower Super Output Area of residence); existing diagnoses of health care conditions; and numbers of prescription medications. Many of these variables feature in the Combined Predictive Model (Wennberg et al. 2006), but are included in our matching algorithm so that the groups will be similar in terms of these more detailed variables as well as the overall risk score.

To increase the number of potential controls available for the analysis, we will assign each potential control up to 12 index dates, one for each month from January 2014 to December 2014, assuming that the patient is alive at each of these time points. Baseline variables (*e.g.*, age and prior primary and secondary care utilisation) will then be calculated at each of these index dates, and each control patient will provide the matching algorithm with up to twelve observations.

Matched control observations will be selected using the genetic matching algorithm, which is a computer-intensive search procedure that produces more closely balanced groups than traditional approaches such as nearest neighbour matching on the propensity score (Sekhon & Grieve 2012). Balance will be assessed using the standardised difference, which is defined as the difference in means as a proportion of the pooled standard deviation (Austin 2009). Although the standardised difference would ideally be minimised without limit, a threshold of 10% has been used to describe meaningful difference (Normand et al. 2001). We will adjust the parameters in the genetic matching algorithm until we obtain standardised differences below these thresholds.

The matching will be conducted with replacement, meaning that a patient might be used as the control for multiple integrated care patients. Our preference is to select two control observations per patient, as this would increase the precision of our estimates of the effect of the integrated care pathway on patient outcomes (*i.e.*, reduce the width of our confidence intervals, meaning we can be more sure about the effect size). However, if it is not possible to obtain adequate balance when selecting two controls per patient, we will rerun the matching algorithm when selecting one control per patient.

Statistical method

After adequate balance has been obtained, we will estimate the effect of the integrated care pathway on patient outcomes by fitting regression models to the data. These models will adjust for residual differences remaining between the groups in the baseline variables after matching. We will include all patients in these analyses, even if they died or moved away from the WELC area during the 12-month follow-up period.

For our primary endpoint, number of emergency hospital admissions, use of Poisson regression seems appropriate, but depending on the empirical distribution of this endpoint, alternative distributions will be considered (*e.g.*, negative binomial). We assume that ordinary least squares (linear) regression will be used

for our secondary endpoint, the number of hospital bed days. On receipt of the line-level data, we will test the suitability of these assumptions, and consider alternative distributions if appropriate. We will follow a similar procedure for other endpoints making sure that assumptions on distributions are appropriate.

Subgroup analysis

The main analysis will be pooled across all three CCGs and across all risk bands, as this approach is necessary to ensure adequate sample size. However, we will also examine whether there is evidence that the integrated care pathway affected the outcomes of certain subgroups of patients more or less than average. Specifically, we will examine subgroups defined by CCG (Newham, Tower Hamlets, and Waltham Forest) and risk band (top 0.5%, 0.5-5%, and 5-20%). We will also distinguish between patients with and without a recorded mental health condition.

The subgroup analysis will be performed using interaction tests, and interpreted using credibility criteria (Sun et al. 2012). To determine whether subgroups were differentially affected by the intervention, we will compare confidence intervals from the interaction tests with the estimated effect from the pooled analysis.

Secondary analysis will also include the evaluation of health care utilisation for patients at the end of life, and evaluation of the whole cohort but accounting for attrition due to patients leaving the programme, dying, or moving out of the WELC area. Attrition will be dealt with using appropriate statistical methods, effectively using a 'person-years' denominator (this can be operationalised by using an offset term in the Poisson regression).

Sensitivity analysis for unobserved confounding

One of the main threats to the validity of this study is unobserved confounding. That is, although we anticipate that the intervention and matched control groups will be similar in terms of observed variables (such as age and prior number of hospital admissions), there may be differences between these groups that we do not observe (such as social isolation). Our findings might be biased if, for example, the patients recruited into the integrated care pathway were disproportionately isolated, and these patients were expected to have worse outcomes than other patients. We note that such selection effects would need to occur beyond the formal eligibility criteria for the integrated care programme, because the formal criteria should be exactly reproducible using the routine data (predictive risk score). However, health care professionals might have had a tendency to enrol patients with some characteristics more than others, and likewise some patients might have been more or less willing to consent to be enrolled than others.

Although there is no definitive way to assess the effect of unobserved confounding, we can compare the mortality rates of the integrated care and matched control groups (West & Thoemmes 2010). On the assumption that the integrated care pathway is unlikely to have had a large positive or negative impact on

mortality within a 12-month period, then differences in mortality rates would make us doubt the performance of the matching. For example, if enrolled patients died at a higher rate than matched control patients, this might suggest that they were in worse health than controls at the point of enrolment (Steventon, Bardsley, Billings, Georghiou, et al. 2012).

A second strategy is to test the robustness of the study findings to hypothetical unobserved confounders that we will simulate (Carnegie et al. 2013). The crucial dimensions are the strength of association between the confounder and treatment assignment, and likewise the strength of the association between the confounder with outcome (for example, if the hypothetical confounder is social isolation, then the degree to which socially isolated individuals are disproportionately enrolled onto the integrated care pathway and the extent to which they have worse outcomes than other patients who are not as socially isolated). We will make various assumptions about these associations and assess how strong they must be to overturn our initial evaluation findings. We will then compare these associations with those observed for known variables (*e.g.*, age) to assess how likely they are to have occurred.

GOVERNANCE

Since WELC has already established an evaluation steering group for the programme, we propose that the scope of this group is expanded to include consideration of this additional work. Formal ethical review is not required for the analyses outlined in the current document, as they constitute a service evaluation and retrospective secondary use of pseudonymised data.

TIMESCALE

The following table shows the anticipated timeframe for the completion of these analyses. This assumes timely approval of the protocol by the evaluation steering group and timely transfer of the line-level data.

Milestone	**Date**
Approval of protocol by the evaluation steering group	June 2015
Permission to access Tower Hamlets data	July 2015
Data transferred	April 2016
Permission to access Newham data, and subsequent additional data transfer	May 2016 (if successful in obtaining permissions)
Preliminary analysis complete	end of October 2016
Present findings to stakeholders	October 2016
Final analysis complete. Report released to the sites	December 2016
Related journal articles submitted for peer-review	March 2016

DISSEMINATION

We will share the evaluation findings with the evaluation steering group at the first available opportunity and use the group's expertise to explore possible explanations for the findings. We also expect to conduct limited further analysis following this conversation. We will submit the final findings to an academic peer-reviewed journal, after seeking comment from the group on the draft manuscript.

HANDOVER

The WELC integrated pioneer programme is already monitoring performance against key metrics using the data cube, while the researcher-in-residence programme is delivering a process-oriented, qualitative and formative evaluation of the implementation of integrated care across WELC. The proposed analysis described above would provide some initial information about the effect of the new integrated care pathway on the utilisation of primary and secondary care and associated costs, as assessed against a matched control group. However, even after the completion of this analysis, there will remain some significant gaps in range of quantitative analyses being planned:

1. Analysis of the effect of the pathway on patient experience and health outcomes;
2. Regular monitoring of the effectiveness of the integrated care pathway; and
3. Information to enable the optimisation of the programme.

Below we outline some thoughts at this stage about these three gaps.

1: Analysis of the effect of the pathway

Analysis of patient experience and outcomes may require more data collection, which could be envisaged as part of the programme (routine data collection for all patients to inform the approach to delivering tailored care) or a research exercise (surveys with defined sampling frames).

2. Regular updates against matched controls

The work described in this document will establish the feasibility of estimating the effect of the integrated care pathway against a matched control group. It would produce some computer code that formats the line-level data sets, produce matched controls, and estimates the effect of the pathway on the utilisation of primary and secondary care and associated costs. Having developed this code, it will be relatively easy to reuse it, for example on a monthly or quarterly basis. Although the approach will depend on local resources, it seems to us that there are some advantages for the Commissioning Support Unit taking this on, since they are close to the data and perhaps could turn around the analysis rapidly, while incorporating the findings into the existing information infrastructure surrounding the data cube. If information on patient experience and outcomes were collected routinely for all patients (as opposed to as a research exercise), then it would be possible to include these outcomes in the matched control analyses.

The Health Foundation would be happy to work with local partners to embed the use of this code, for example by providing:

1. A standardised set of computer programmes to do the analyses, with documentation.
2. 'How-to' guides about matching.
3. Time to address any teething problems with the handover.

We are already exploring whether the NHS could establish a series of 'data labs' to conduct this analysis on a regular basis. In exchange for providing this support, we would appreciate feedback to enable us to fine-tune the computer programmes and how-to guides.

3: Information to enable the optimisation of the programme

Regular matched control analyses are a promising adjunct to the data cube, since they provide robust information of the impact of the pathway. By repeating the analysis on a regular basis, it will be possible to establish whether effectiveness is improving over time. It will also be possible to 'drill down' into the data and understand whether effects are better or worse for subgroups of patients, for example based on risk score or CCG. However, other tools are needed to develop strategies for optimising the performance of the programme. For example, it will be necessary to:

1. Triangulate the findings from the matched control analysis with those of the researcher-in-residence;
2. Understand the reasons for variation in effectiveness between subgroups; and
3. Understand the contribution made by each of the interventions that are delivered to patients on the pathway.

Compared to the previous strand of activity, this seems less about establishing new information infrastructure, as the questions of interest will change over time. Specialist analytical and research skills may be needed.

DATA SPECIFICATION

Following approval of this protocol, we hope to work with the local Commissioning Support Unit (CSU) to develop a detailed specification for the line-level data. These data sets will then need to be assembled and transferred to the secure infrastructure at the Health Foundation for analysis.

Patient-level data for this project will be held and processed on the secure infrastructure hosted by the Health Foundation at 90 Long Acre in London. All data within this environment are held in encrypted form. Access to the processing servers is provided through thin clients (*i.e.*, terminals that can only be used to connect to the secure environment and have no further connectivity such as email or internet), whereby all communications between the client and the secure environment are encrypted (key strokes, mouse clicks and screen refreshes only). Research outputs (*e.g.*, regression coefficients) are manually

checked before release for publication to ensure the research findings are non-disclosive. All staff members with access to the data are trained in information security, data protection legislation and statistical disclosure control following best practice guidelines, and are contractually held to non-disclosure agreements at the risk of disciplinary action and/or dismissal.

We will require all data sets to be pseudonymised before transfer; no patient identifiable information (*e.g.*, name, address, full date of birth) are required to complete the analysis. Pseudonymising NHS number (*i.e.*, apply a one-way has algorithm to the NHS number, only known to the data controller) allows the data to be updated and extended, if necessary. The data sets will need to include line-level data for the top twenty per cent of patients at risk of emergency hospital admission aged over fifty years in the WELC area[1], spanning the period 1 January 2011 to 31 December 2015. Both age and risk-score vary over time, so the exact definition of these extracts will be defined in the detailed specification. We will need Read-coded GP data on registrations with GPs, prescriptions, diagnoses, and encounters, together with line-level data from the admitted patient care, outpatient and accident and emergency commissioning data sets. It would greatly assist us if some derived variables were available for the analysis (for example, the Combined Model score and constituent variables) – this would also help us standardise our variable definitions with those used locally.

[1] Note that only data from Tower Hamlets CCG and Newham CCG is to be included.

References

Austin, P.C., 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28, pp.3083–3107.

Carnegie, N.B., Harada, M. & Hill, J., 2013. Assessing sensitivity to unmeasured confounding using simulated potential confounders. *Working paper*.

Kaplan, E.L. & Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), pp.457–481.

Normand, S.-L.T. et al., 2001. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of Clinical Epidemiology*, 54(4), pp.387–398. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11297888.

Roland, M. et al., 2005. Follow up of people aged 65 and over with a history of emergency admissions: analysis of routine admission data. *BMJ*, 330, pp.289–292. Available at: http://www.bmj.com/content/330/7486/289.short [Accessed August 21, 2012].

Sekhon, J.S. & Grieve, R.D., 2012. A matching method for improving covariate balance in cost-effectiveness analyses. *Health Economics*, 21(6), pp.695–714. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21633989 [Accessed March 2, 2013].

Steventon, A., Bardsley, M., Billings, J., Dixon, J., et al., 2012. Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. *BMJ*, 344, p.e3874. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3381047&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2012].

Steventon, A., Bardsley, M., Billings, J., Georghiou, T., et al., 2012. The role of matched controls in building an evidence base for hospital-avoidance schemes: a retrospective evaluation. *Health Services Research*, 47(4), pp.1679–1698. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22224902 [Accessed August 15, 2012].

Sun, X. et al., 2012. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ*, 344(mar15 1), pp.e1553–e1553. Available at: http://www.bmj.com/content/344/bmj.e1553 [Accessed March 13, 2013].

Wennberg, D. et al., 2006. *Combined predictive model: final report and technical documentation*, London: King's Fund.

West, S.G. & Thoemmes, F., 2010. Campbell's and Rubin's perspectives on causal inference. *Psychological methods*, 15(1), pp.18–37. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20230100 [Accessed August 16, 2012].