Electronic Medical Record-Assisted Design of a Cluster-Randomized Trial to Improve Diabetes Care and Outcomes

Thomas E. Love, PhD^{1,2}, Randall D. Cebul, MD^{1,2,3}, Douglas Einstadter, MD, MPH^{1,2,3}, Anil K. Jain, MD⁴, Holly Miller, MD, MBA⁵, C. Martin Harris, MD, MBA⁶, Peter J. Greco, MD², Scott S. Husak, BS¹, and Neal V. Dawson, MD^{1,2,3}, for the DIG-IT Investigators

¹Center for Health Care Research and Policy, Case Western Reserve University-MetroHealth Medical Center, Cleveland, OH, USA; ²Department of Medicine, School of Medicine, Case Western Reserve University-MetroHealth Medical Center, Cleveland, OH, USA; ³Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; ⁴Information Technology Division and Department of General Internal Medicine, Medicine Institute, Cleveland Clinic Foundation, Cleveland, OH, USA; ⁵University Hospitals, Cleveland, OH, USA; ⁶Cleveland Clinic Foundation, Cleveland, OH, USA.

BACKGROUND: Electronic medical records (EMRs) have the potential to facilitate the design of large cluster-randomized trials (CRTs).

OBJECTIVE: To describe the design of a CRT of clinical decision support to improve diabetes care and outcomes.

METHODS: In the Diabetes Improvement Group-Intervention Trial (DIG-IT), we identified and balanced pre-assignment characteristics of 12,675 diabetic patients cared for by 147 physicians in 24 practices of 2 systems using the same vendor's EMR. EMR-facilitated disease management was system A's experimental intervention; system B interventions involved patient empowerment, with or without disease management. For our sample, we: (1) identified characteristics associated with response to interventions or outcomes; (2) summarized feasible partitions of 10 system A practices (2 groups) and 14 system B practices (3 groups) using intra-cluster correlation coefficients (ICCs) and standardized differences; (3) selected (blinded) partitions to effectively balance the characteristics; and (4) randomly assigned groups of practices to interventions.

RESULTS: In System A, 4,306 patients, were assigned to 2 groups of practices; 8,369 patients in system B were assigned to 3 groups of practices. Nearly all baseline outcome variables and covariates were well-balanced, including several not included in the initial design. DIG-IT's balance was superior to alternative partitions based on volume, geography or demographics alone.

CONCLUSIONS: EMRs facilitated rigorous CRT design by identifying large numbers of patients with diabetes and enabling fair comparisons through preassignment balancing of practice sites. Our methods can be replicated in other settings and for other conditions, enhancing the power of other translational investigations.

KEY WORDS: cluster-randomized trial; study design; health information technology; clinical decision support; diabetes mellitus.

J Gen Intern Med 23(4):383–91
DOI: 10.1007/s11606-007-0454-3
© Society of General Internal Medicine 2007

INTRODUCTION

Diabetes is a complex, chronic, and costly disease, affecting nearly 21 million Americans at an annual cost of \$132 billion.¹ Major gaps are apparent between what may be achieved and current performance, including gaps in glycemic, blood pressure, and lipid control, inadequate risk-factor monitoring and treatment, and failure to provide appropriate preventive services.² The Centers for Disease Control and Prevention estimate that adequate attention to these measures would decrease cardiovascular disease mortality by 30%, blindness by 90%, and incidence of end-stage renal disease by 50%.¹

Translating our understanding of diabetes epidemiology into effective interventions has been challenging. Modern approaches to improving diabetes care often include decision support for physicians and/or patients.³ Several recent investigations employed advanced health information technologies and clinical decision support systems using functions enabled by electronic medical record (EMR) systems.^{4–6} A meta-analysis of quality improvement strategies for Type 2 diabetes³ reported that while most approaches produced small to modest improvements, those involving team changes and case management appeared most promising.

The preferred design to test interventions intended to change physician and/or patient behavior is the cluster randomized trial (CRT).^{7,8} In CRTs, the unit of randomization (typically a group, such as physicians in a clinical practice) is not necessarily the analytic unit of interest (typically a patient). Well-designed CRTs often capitalize upon the effects of within-group learning, such as that which occurs with clinical decision support systems and team approaches to care, while minimizing cross-group contamination. Published CRTs often fail to adhere to rigorous design and analytic standards, as indicated in a systematic review of 152 primary care CRT publications

from 1997–2000.^{9,10} Design flaws often relate to the failure to minimize baseline cross-group differences in important prognostic characteristics,⁹ resulting in biased comparisons. Common analytic flaws include failure to account for cross-group differences in power estimates or ignoring clustering effects.

While EMR functions to provide decision support have been widely recognized,¹¹ 2 EMR attributes that facilitate rigorous CRT design have received little attention. EMRs enable the identification of virtually all patients with a given condition who receive care within a health care system, by particular physicians, and at specific practice sites. EMRs can also provide detailed pre-intervention data to enable the balancing of groups of patients or practices on important clinical, sociodemographic, and behavioral characteristics, allowing the investigator to build study groups which yield fair comparisons.

EMRs provide detailed and prognostically relevant data upon which to balance study groups and assess the power of a CRT. Without an EMR, researchers designing CRTs most often use published intra-cluster correlation coefficient (ICC) estimates to estimate sample size requirements.^{4,7,8,12–14} Methods for matching or stratifying to produce study groups have involved sampling for relevant features, surveying historical outcomes before assignment,^{15–17} or assigning study groups to balance available features, such as geographical location,^{5,18,19} patient or physician volume,^{5,18} practice type,²⁰ or sociodemographics.²¹

Here, we describe the design of the Diabetes Improvement Group-Intervention Trial (DIG-IT), a CRT of EMR-catalyzed decision support to improve the care and outcomes of adult diabetic patients. We highlight the use of the EMR to facilitate identification and pre-randomization balancing on detailed characteristics of 12,675 diabetic patients cared for by 147 physicians in 24 practices of 2 health systems using the same commercial EMR system. We also compare the performance of our EMR-assisted design to several alternatives without EMR information.

METHODS

Diabetes Improvement Group-Intervention Trial

DIG-IT aims to estimate the incremental effects of EMR-facilitated disease management, patient empowerment, or the combination, compared to unaltered EMR (“usual care”), on measures of: (1) patient care quality and safety; and (2) health services utilization. As shown in Figure 1, System A’s 2 groups include: (1) usual clinical care using the EMR with standard features, and (2) EMR care amplified by EMR-centered disease management for diabetes mellitus [DM²]. System B’s 3 groups

are: (3) EMR care amplified by an opportunity for patients to access components of their records through a web portal (MyChart®) [MC], (4) MC care amplified by opportunities for patients to use diabetes-specific functionalities to aid self-management [MC PLUS], and (5) EMR care amplified by DM² and MC PLUS, in combination [BOTH].

Key outcomes include changes in hemoglobin A1c values over the 2-year trial and a summary score—the proportion met of 8 measures based on American Diabetes Association recommendations,²² each linked to an EMR-facilitated alert in our DM² intervention. These include: achievement of optimal levels of hemoglobin A1c ($\leq 7\%$), LDL cholesterol (≤ 100 mg/dl), systolic blood pressure (≤ 130 mmHg), and body mass index (≤ 30 kg/m²); annual ophthalmologic examination referral, pneumococcal immunization, nonsmoking status, and appropriate treatment of microalbuminuria with angiotensin converting enzyme (ACE) inhibitors or angiotensin receptor blocking (ARB) agents. We study utilization both overall and characterized as undesirable (hospitalizations, emergency visits), desirable (use of indicated medications and tests; primary care visits, immunizations, ophthalmologic referrals), or other.

Setting. DIG-IT includes physicians caring for diabetic patients in 24 outpatient primary care practices of 2 academic health care systems in northeast Ohio. System A is a public system with 10 primary care practices located principally in underserved areas in Greater Cleveland; system B is a private not-for-profit system with 14 practices in urban and suburban locations. Since 2000, each system has implemented the same commercial EMR (Epic® Systems, Madison, Wisconsin.) In 2004, system B added “tethered” personal health record functionalities so that patients could access selected medical record components.

Subjects. The analyses reported here include diabetic patients at least 18 years of age regularly seen by primary care physicians (specialists in General Internal Medicine, Family Practice, and Medicine-Pediatrics) in 24 practices. “Regularly seen” patients were seen at least once by the same physician in an eligible primary care clinic in each of the 2 years (2003–2004) preceding DIG-IT. Diabetes is identified by its appearance on the problem list or by a diabetic medication in the medications list and at least 1 encounter in 2003–2004 identifying diabetes as the reason for the visit. In system A, we excluded a small practice comprised exclusively of Asian-American patients and a teaching practice where residents serve a major role in clinical care. In DIG-IT, patients will be excluded if they cannot speak English, are blind or pregnant, or, in system B, if they state that

Figure 1. Schematic of DIG-IT design showing the 5 arms, within 2 health systems. *EMR* Usual clinical care using the electronic medical record with standard features, *DM²* EMR amplified by EMR-centered disease management for diabetes mellitus, *MC* EMR amplified patient opportunity to access records through MyChart® web portal, *MC PLUS* MC amplified by patient opportunity to use added diabetes-specific functionalities to aid self-management, *BOTH* DM² and MC PLUS, in combination.

they are physically incapable of using a computer. DIG-IT places no constraints on usual patient or physician decision making, and no decrement of care associated with study methods is expected. Each system's Institutional Review Board approved all relevant DIG-IT protocols.

Assigning Practices to Study Arms: the Preassignment Balancing Process

We outline our procedure for partitioning practices into study groups to balance detailed baseline characteristics within each system. To begin, clinician-investigators identified clinical (e.g., A1_c level and trend), adherence-related (e.g., “no show” rates, smoking and vaccination status), and demographic factors likely to be associated with patients' response to disease management and/or changes in A1_c levels. From the EMR, we gathered data for encounters and laboratory results in the 2 years preceding DIG-IT for all regularly seen adult diabetes patients. Tables 2 and 3 identify our “balancing set”, including proportions of patients who were female, African American, of Hispanic ethnicity; prescription for insulin at any time; record of a pneumonia vaccination; documented flu vaccination, emergency department visit or “no show” for a PCP visit in the last 12 months; and last recorded systolic blood pressure and hemoglobin A1_c values. For patients with at least 2 A1_c values separated by 30 days or more, we estimated practice-level mean annualized trends in glycemic control over time.

Next, we summarized the balancing set characteristics for alternative partitions (potential assignments of practices to study groups) within each system. In system A, we studied 336 unique partitions of the 10 practices, including the 126 possible partitions into 2 study groups containing 5 practices each and the 210 partitions which create 1 group of 4 and 1 group of 6 practices. In system B, practical considerations required preassignment of 4 practices (2,069 patients) to MC and combining 2 practice pairs which share clinicians. For practices not preassigned to MC, we enumerated 72 partitions of 6 practices and 2 sets of “coupled” practices to the MC PLUS and BOTH study groups, representing all alternatives where each study group contains 4 to 6 practices.

Rather than adopt a purely statistical approach to optimize balance,^{23,24} the investigative team met to consider 10 candidate partitions of system A practices and 9 candidates for system B, in each case, blinded to practice identification. The lead author calculated absolute standardized differences (difference in means as a percentage of the pooled SD)^{25–27} across each characteristic to inform the selection of candidates. Equipped with descriptive summaries of each balancing set characteristic for each candidate partition, the team discussed which alternatives reflected the “best” balance, and then selected a partitioning of the practices. Finally, these were randomly assigned to study groups by coin toss.

The most important balance measure for a CRT is the intra-cluster correlation coefficient (ICC), the ratio of the between-clusters variance to the total variance, which takes a value between 0 and 1.²⁸ The ICC for a trial's key outcomes at baseline affects both the effective sample size and required analytic approach, with larger ICCs having greater impact. Careful ICC estimation is crucial.^{9,13,14,29–32} Large ICCs occur if, for instance, there is consistent management of patients

Table 1. Baseline Characteristics of Diabetic Patients in Practice Sites in Systems A and B

Baseline Characteristics	System A	Range for System A Sites	System B	Range for System B Sites
Number of practices	10	–	14	–
Number of diabetic patients	4,306	77–1,079	8,369	220 – 1,239
Number of primary care Physicians	73	4–17	74	5–22
Sociodemographics				
Age in years, mean	58.5	56.9–60.0	63.1	60.5–66.7
Female (%)	65.0	52.2–76.9	51.1	42.8–59.7
African American (%)	48.9	4.5–98.2	23.0	0.7–75.6
Hispanic ethnicity (%)	10.1	0–22.6	1.2	0–10.2
Insurance status				
Commercial (%)	29.1	15.4–53.1	55.9	40.8–64.0
Medicaid (%)	18.0	2.9–35.1	3.3	1.0–12.6
Medicare (%)	40.2	24.8–53.6	39.0	32.1–54.0
Uninsured/self-pay (%)	12.6	6.8–23.4	1.8	1.1–3.2
Neighborhood income estimate: mean (in ,000s)	31.9	23.1–57.9	47.7	32.4–63.1
Neighborhood education estimate: percent with less than high school diploma	29.4	11.8–38.1	15.6	10.5–26.0
Glycemic control				
Annualized A1c trend, slope	–0.62	–0.89–0.48	–0.46	–1.00–0.29
Most recent A1c, mean	7.47	7.14–8.04	7.19	6.97–7.48
Most recent A1c ≥ 9.0 (%)	17.8	12.7–26.0	10.8	6.2–15.2
Summary score elements				
Most recent A1c ≤ 7.0 (%)	51.7	37.5–56.6	56.7	49.8–64.4
Most recent BMI, mean	34.4	33.0–35.2	32.6	32.0–33.3
Most recent systolic BP, mean	136.2	134.9–139.2	132.8	128.4–141.2
Most recent LDL, mean	116.6	113.8–119.7	99.1	94.4–109.4
Current smoking (%)	23.9	17.7–33.9	10.8	7.0–17.7
Proteinuria on ACE/ARB (%)	80.7	71.3–86.5	80.1	73.5–85.7
Eye examination in 2004 (%)	35.0	27.0–38.3	32.0	6.7–62.3
Pneumococcal vaccination (%)	48.1	37.0–68.8	55.3	47.4–70.5
Other characteristics				
Influenza vaccination in 2004 (%)	32.4	28.6–35.3	44.1	35.2–56.1
Prescription for insulin (%)	19.0	15.7–22.1	21.2	13.9–31.9
≥ 1 ED visit in the previous year (%)	23.4	9.5–31.0	4.7	0–19.1
≥ 1 “no show” for PCP visit in the previous year (%)	45.3	27.2–60.2	17.3	10.0–28.6

Table 2. Balance Achieved by DIG-IT for System A (Variables Used in Balancing Process)

Variable	System A Study Arms		ICC*	SzDiff (%) [†]
	DM ²	EMR		
Number of practices	5	5	–	–
Number of diabetic patients	2281	2025	–	–
Female (%)	67.1	62.7	0.004	9
African American (%)	48.7	49.1	< 0.001	–1
Hispanic ethnicity (%)	9.7	10.5	< 0.001	–3
Current smoking (%) [‡]	25.2	22.6	0.001	6
Most recent A1c ≤7.0 (%) [‡]	49.0	54.6	0.006	–11
Most recent A1c ≥9.0 (%)	18.7	16.9	0.001	5
Prescription for insulin (%)	18.5	19.6	< 0.001	–3
≥1 ED visit in last year (%)	26.9	19.4	0.015	18
≥1 “no show” in last year (%)	43.1	47.8	0.004	–9
Pneumococcal vaccination (%) [‡]	43.9	52.8	0.015	–18
Influenza vaccination in 2004 (%)	33.0	31.7	< 0.001	3
Annualized A1c trend, mean (SD)	–0.66 (2.2)	–0.57 (2.4)	< 0.001	–4
Most recent A1c, mean (SD)	7.3 (1.8)	7.6 (2.0)	0.014	–17
Most recent systolic BP, [‡] mean (SD)	136 (21)	136 (21)	< 0.001	–1

*ICC Intra-cluster correlation coefficient

[†]SzDiff Standardized difference, as a percentage of the pooled SD[‡]A binary indicator of performance on this measure is a part of the 8-element summary score planned as a key DIG-IT outcome.

within each practice, but substantial variation across practices.^{28,33} Reported ICCs for primary care processes often fall between 0.05 and 0.15, while those for patient outcomes are generally below 0.05.^{33,34}

The primary impact of statistical dependence implied by a positive ICC is that the cluster-randomized trial loses power relative to a trial randomized at the patient level.^{7,8,35} The CRT's effective sample size for an outcome is inversely related to that outcome's ICC. To obtain an effective sample size for an outcome in terms of the required number of patients for a trial randomized at the patient level, we deflate the CRT's sample size by a design effect, $DE = 1 + (\bar{n} - 1)ICC$, which depends on the average cluster size $\bar{n}$ and the estimated ICC.^{8,33} Using our historical EMR data, we estimated ICCs for each characteristic for each partition using STATA's *loneway* procedure,³⁶ enabling us to directly compare the selected DIG-IT study groups to all alternative partitions.

To assess balance more thoroughly, we estimated standardized differences for our selected DIG-IT partition to compare the 2 system A study groups, and the 2 system B groups (MC PLUS and BOTH), where allocation was not predetermined. Standardized differences below 10% in absolute value indicate well-balanced distributions.²⁵ Finally, we examined DIG-IT's balance for several variables not included in the balancing set, including geocoded estimates of neighborhood income and educational levels drawn from the U.S. Census, at the census block level for system A, and the zip code level for system B,

with educational estimates in each case specific to the patient's race and gender.

RESULTS

Table 1 describes preassignment central tendencies and ranges of key characteristics of the studied practices. Striking differences exist both within and across systems. For instance, while 48.9% of diabetic patients in system A are African American, system's A 10 practices range from 4.5% to 98.2% African American. In system B, 23% of diabetics are African American overall, but individual practices range from 0.7% to 75.6% African American. Table 1 also displays substantial between- and within-system variation in insurance type, geocoded race- and gender-matched census estimates of income and educational attainment, baseline glycemic control, ADA summary components and measures of undesirable utilization.

Table 3. Balance Achieved by DIG-IT for System B (Variables Used in Balancing Process)

Variable	System B Study Arms			ICC*	SzDiff (%) [†]
	MC	MC PLUS	BOTH		
Number of practices	4	6	4	–	–
Number of diabetic patients	2069	3115	3185	–	–
Female (%)	49.3	52.9	50.6	0.001	5
African American (%)	21.4	19.2	27.7	0.012	–20
Hispanic ethnicity (%)	0.8	2.4	0.4	0.010	18
Current smoking (%) [‡]	10.7	10.3	11.2	< 0.001	–3
Most recent A1c ≤7.0 (%) [‡]	59.0	57.6	54.5	0.002	6
Most recent A1c ≥9.0 (%)	9.7	10.1	12.2	0.002	–7
Prescription for insulin (%)	19.0	21.6	22.3	0.001	–2
≥1 ED visit in last year (%)	4.5	3.7	5.7	0.002	–10
≥1 “no show” in last year (%)	15.8	18.7	16.8	0.001	5
Pneumococcal vaccination (%) [‡]	56.7	57.2	52.6	0.002	9
Influenza vaccination in 2004 (%)	46.2	44.7	42.2	0.001	5
Annualized A1c trend, mean (SD)	–0.45 (1.8)	–0.48 (2.1)	–0.43 (2.2)	< 0.001	–2
Most recent A1c, mean (SD)	7.1 (1.4)	7.2 (1.4)	7.3 (1.5)	0.002	–8
Most recent systolic BP, [‡] mean (SD)	133 (18)	133 (18)	132 (18)	< 0.001	4

*ICC Intra-cluster correlation coefficient across all 3 groups

[†]SzDiff Standardized difference comparing MC PLUS to BOTH, as a percentage of the pooled SD[‡]A binary indicator of performance on this measure is a part of the 8-element summary score planned as a key DIG-IT outcome.

In contrast, Tables 2 and 3 highlight impressive similarities between study groups after the balancing process. In Table 2, system A's selected partition has ICC values well below 0.01 and absolute standardized differences below 10% for 10 of the 14 characteristics. Table 3 displays the balance in system B, where all characteristics other than race and ethnicity show very small ICCs and standardized differences.

Figure 2 presents boxplots of ICC estimates across all 336 alternative partitions of practices in system A on the balancing set of characteristics (excluding patient volume.) In each plot, the ICC achieved by the partition selected for DIG-IT is indicated by a triangle. The selected DIG-IT study groups appear highly comparable, with solid balance (low ICC) on virtually all variables. Most characteristics show modest ICC variation across the 336 alternative partitions. However, a poor choice of partition could lead to very poor balance on some characteristics, particularly race/ethnicity, but also most recent value of A1c, gender, and rates of insulin prescription and vaccination. Figure 3 provides analogous information for the 72 partitions in system B. Again, the DIG-IT choice yields study groups with small ICCs in absolute terms and in comparison to the available alternatives.

Comparing DIG-IT's Balance with Plausible non-EMR Alternatives

Without EMR information, the assignment of each system's practices to DIG-IT study groups might have been along

geographical lines,^{5,18,19} by balancing characteristics available from non-EMR data sources, using simple demographics,²¹ or patient and physician volume.^{5,18} To help estimate the value of incorporating EMR data into the balancing process, we compare the DIG-IT balance on key variables to that produced by likely non-EMR alternatives. Each option leads to substantially poorer balance than the DIG-IT selection for at least some key outcomes and prognostic variables, often race and volume. A table detailing the balance of each alternative is available from the first author.

Geography. In Greater Cleveland, forming system A's 5 Eastern practices into 1 study group (and 5 Western practices into the other) yields a substantially poorer match than the DIG-IT selection. The racial balance is much worse (24% vs 82% African American, compared to 49% in each DIG-IT group), volume is less balanced (2,482 vs 1,824 patients compared to 2,281 and 2,025), and hemoglobin A1c values are less similar (mean last A1c values are 7.24 and 7.78, compared to 7.32 and 7.64). System B's MC PLUS and BOTH arms would have 4 Western practices and 6 in the East and are also substantially less balanced than the DIG-IT selection in race (5% and 34% African American, compared to 19% and 28%) and volume (2,219 and 4,081 patients compared to 3,115 and 3,185). In each system, the DIG-IT partition shows better balance (i.e., a smaller ICC) than the geographic approach for 11 of the 15 characteristics in Tables 2 and 3, and is never substantially worse.

Figure 2. Balance achieved in system A on characteristics used in design. Boxplots indicate intra-cluster correlation coefficients (ICCs) for each of the 336 available partitions of system A's practices. The DIG-IT choice is marked with a triangle. Note that the maximum ICC (across all 336 possible partitions) for the proportion of African-American patients in system A is 0.677 and that 16 of the 336 arrangements have ICCs exceeding 0.40, and are thus not shown in the Figure, which is scaled to focus on the relevant information for other characteristics.

Figure 3. Balance achieved in system B on characteristics used in design. Boxplots indicate intra-cluster correlation coefficients (ICCs) for each of the 72 studied partitions in system B. The DIG-IT partition is marked with a triangle.

Volume. The selected DIG-IT study groups are closely matched on patient volume. System A shows 2,281 and 2,025 patients and 44 and 29 physicians, respectively, in the 2 arms. Alternative partitions with closer matches on both patient and physician volume are substantially less balanced on the proportion of African-American patients. The system B choice balances patient volume across the MC PLUS ($n=3115$) and BOTH ($n=3185$) arms, and these are also well balanced on number of physicians (55 and 63, respectively.)

Demographics. Without an EMR but using, for instance, an insurance database, we might have sufficient data to select a partition which best balances age and gender. In each system, this age/sex partition produces a result with very poor balance on race and ethnicity and substantial differences in patient volume across the exposures. Restricting the candidates to those partitions which have a difference in patient volume of 500 or less, the selected DIG-IT partition is substantially better on race/ethnicity, sociodemographics (education and income, particularly) and baseline glycemic control. We also considered a situation where race was available in addition to age and gender. Again, restricting to partitions with patient volume differences of 500 or less, in system A, the DIG-IT partition is the best choice to balance age, gender, and race/ethnicity. The system B partition which best balances age, gender, and race/ethnicity is well balanced overall, with the exceptions of neighborhood estimates of educational attainment (17.4% less than high school graduates vs 13.5%) and mean income (\$42,700 vs \$51,870.)

Table 4. Balance Achieved by DIG-IT for Selected Variables NOT Used in Design (System A)

Variable	DM ²	EMR	ICC*
Proteinuria on ACE/ARB (%) [†]	82.6	78.6	0.004
Eye examination in 2004 (%) [†]	34.8	35.2	< 0.001
Most recent BMI, [†] mean (SD)	34.4 (8.9)	34.5 (8.6)	< 0.001
Most recent LDL, [†] mean (SD)	115 (36)	118 (40)	0.003
Number of primary care physicians	44	29	–
Percent commercial insurance (%)	28.8	29.5	< 0.001
Percent Medicaid insurance (%)	19.2	16.7	0.002
Percent Medicare insurance	40.3	40.1	< 0.001
Percent uninsured/self-pay	11.7	13.6	0.001
Age in years, mean (SD)	58.7 (13.3)	58.2 (13.6)	< 0.001
Neighborhood income estimate in 1,000s, mean (SD)	30.7 (14.6)	33.3 (18.2)	0.011
Neighborhood estimate: percent with less than HS diploma	30.2 (20.0)	28.6 (19.8)	0.003

Table 5. Balance Achieved by DIG-IT for Selected Variables NOT Used in Design (System B)

Variable	MC	MC PLUS	BOTH	ICC*
Proteinuria on ACE/ARB (%) [†]	77.6	80.5	81.2	< 0.001
Eye examination in 2004 (%) [†]	33.4	26.3	36.6	0.014
Most recent BMI, mean (SD)	32.6 (7.5)	32.8 (7.7)	32.5 (7.8)	< 0.001
Most recent LDL, mean (SD)	98 (32)	99 (33)	100 (34)	< 0.001
Number of primary care physicians	33	55	63	–
Percent commercial insurance	59.5	53.2	56.2	0.003
Percent Medicaid insurance	2.0	4.6	2.8	0.005
Percent Medicare insurance	36.7	40.3	39.3	0.001
Percent uninsured/self-pay	1.8	1.9	1.7	< 0.001
Age in years, mean (SD)	62.3 (13.1)	62.9 (13.3)	63.8 (13.1)	0.003
Neighborhood income estimate in, 000s, mean (SD)	48.5 (14.9)	46.7 (14.0)	48.2 (18.0)	0.003
Neighborhood estimate: percent with less than HS diploma	16.1 (9.0)	15.2 (9.5)	15.6 (9.9)	0.001

*ICC Intra-cluster correlation coefficient

[†]A binary indicator of performance on this measure is a part of the 8-element summary score of ADA recommendations planned as a key DIG-IT outcome.

Balance of Variables Not Accounted for in Design

Having settled on a final DIG-IT partition by balancing the number of practices, patient volume, and the other variables identified in Tables 2 and 3, we examined the performance on several additional practice-level characteristics not used in the design. Tables 4 and 5 summarize age, 4 data elements related to the summary score outcome, and 3 proxies for socioeconomic status, specifically each practice's distribution of insurance status and geocoded estimates of neighborhood patient income and educational attainment. Although none of these variables were incorporated directly in the balancing process, all but 1 in each system exhibit very small ICCs (below 0.01) in the selected DIG-IT partition.

Effective Sample Size

Cluster-randomized trials studying multiple outcomes will have different statistical power characteristics for each outcome. Here, we illustrate effects associated with slopes of measured hemoglobin A_{1c} levels, designated as the primary outcome during the planning phase. Within system A, our estimated ICC for A_{1c} slope is 0.00024. Assuming our 2 study groups (DM² and EMR) will each have 2,001 patients, the design effect is 1.48, which implies that DIG-IT will have power equivalent to that of a trial ran-

domized at the patient level with approximately 1,352 patients in each arm. In system B, again assuming 2,001 patients per study group, the estimated design effect is 1.025, implying that DIG-IT will have an effective sample size of 1,952 patients in each of the 3 system B arms. Our design thus yields approximately 90% power to detect (with a 2-sided $\alpha=0.05$ test) a difference in A_{1c} change as small as 0.25% across the 2 study arms in system A or across any pair of the 3 system B arms.

DISCUSSION

To our knowledge, this report describes the first cluster randomized trial to use data from electronic medical records to explicitly balance important characteristics before assigning practices to study groups and assigning study groups randomly to interventions. The trial was designed both to optimize fair comparisons across groups and to capitalize on within-practice learning and information sharing that accompanies the introduction of clinical decision support tools. By explicitly balancing practice characteristics within study groups before assignment, we were able to minimize baseline differences, thus reducing substantially the potential for selection bias,³⁷ the need for extensive covariate adjustment in our final models, and the impact on effective sample size.^{7,8,23} The relatively small effective sample size observed in many CRIs³ implies substantial imbalance between intervention groups on important baseline characteristics.⁸ Indeed, through a combination of straightforward identification of diabetic patients and pre-study balancing, the effective sample size of the current investigation for glycemic control exceeds all published cluster randomized trials reported in a recent meta-regression analysis.³ In that report, baseline differences in study groups of 1 investigation³⁸ led the authors to estimate that the reported sample size of 3,608 subjects yielded an effective sample size of only 1,269 subjects, markedly reducing power to detect intervention effects on glycemia.³ Naturally, our effective sample size will be reduced when studying subgroups and outcomes with larger associated ICCs.

As reflected by very small intra-cluster correlation coefficients (ICCs), Tables 2 and 3 highlight the similarities across study groups among variables selected a priori as likely to have strong prognostic relationships either to patient adherence, such as appointment show rates and current smoking, or to intermediate outcomes, such as systolic blood pressure levels and 2-year trends in hemoglobin A_{1c} values. In addition, there was substantial similarity across groups among variables that were not explicitly part of the balancing procedure (Tables 4 and 5). These include such important sociodemographic features as insurance status, estimated neighborhood income, and educational attainment,^{39,40} process of care measures such as appropriate pharmacologic attention to microalbuminuria, and recent ophthalmologic examinations; and non-glycemia-related intermediate outcome measures, such as body mass index and LDL cholesterol values.

The DIG-IT design is not without limitations. In our balancing process, we identified significant cross-group differences in many important parameters between the practices in the 2 health care systems (Table 1). Although we initially intended to undertake a single trial with study groups formed across systems, these differences necessitated a design focused on within-system analyses and interventions, including

1 shared component across systems that featured physician-centered clinical decision support (the principal intervention in system A.) Nonetheless, our attention to baseline data enabled us to avoid unbalanced comparisons. Second, while we had the latitude to assign most practices to study groups on the basis of their contributions to overall balance, and to assign study groups randomly to interventions—hard-won latitude based on negotiations and the good will of practice leaders in the 2 systems—logistical constraints necessitated combining 2 pairs of practices in system B (due to “itinerant” physicians) and preassigning 4 system B practices to the MC arm. Such constraints and realities are commonplace in most cluster-randomized trials and are best recognized at the design stage, where their effects can be minimized. Third, as “open” health care systems in which patients are at liberty to seek aspects of their care elsewhere, we are aware that the data used in our balancing may be incomplete. For example, we suspect that data are incomplete pertaining to eye examinations, which may be differentially available at different geographic sites, and to emergency department visits, which due to urgency, may occur at hospitals closer to home than patients’ ambulatory care site and main campus hospital. Finally, we recognize in our pre-intervention analyses the presence of certain residual imbalances across study groups, including “last A1c” values in system A and the proportion of African-American patients in system B. Should these potentially important differences be observed during the study period, we need to recognize them and conduct necessary statistical adjustments.⁴¹

Comment. Cluster-randomized trials are increasingly common and there are appropriate approaches for examining the effects of changes in systems of care, often directed at group practices or practicing physicians, on quality of care and patient outcomes. Yet because “birds of a feather flock together,” subjects within a group frequently are more similar to one another than they are to subjects of different groups, confounding cross-group comparisons of intervention effects. Balancing of practice groups is thus necessary to make fair cross-group comparisons and simultaneously enable within-practice learning and sharing of information about the new systems. The increasing adoption of electronic medical record systems, especially among large health care systems with multiple geographically distinct practice sites, should facilitate better designs and analyses of cluster-randomized trials.

Acknowledgments: An earlier version of this work was presented at the Society for Medical Decision Making Annual Meeting in Boston, October 2006. This work is supported in part by grant no. R01 HS15123 from the Agency for Healthcare Research and Quality. The authors wish to thank Mary Jo Roach, Ph.D., and Mark Votruba, Ph.D. for their substantial contributions to DIG-IT.

Conflicts of Interest: Dr. Love has received an honorarium for a short course given to Amgen, Inc. Dr. Einstadter has served as a consultant to Medical Mutual of Ohio. Dr. Dawson has served as a consultant to Gerson Lehrman Group. None of these activities are related to the research described in this report.

Corresponding Author: Thomas E. Love, PhD; Center for Health Care Research and Policy, Case Western Reserve University-MetroHealth Medical Center, 2500 MetroHealth Drive, Cleveland, OH 44109-1998, USA (e-mail: Thomas.Love@case.edu).

REFERENCES

- Centers for Disease Control and Prevention. National Diabetes fact sheet: general information and national estimates on diabetes in the United States, 2003 Rev. ed. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2004.
- Saaddine JB, Cadwell B, Gregg EW, Engelgau MM, Vinicor F, Imperatore G, et al. Improvements in diabetes processes of care and intermediate outcomes: United States, 1988–2002. *Ann Intern Med.* 2006;144(7):465–74, Apr 4.
- Shojania KG, Ranji S, McDonald KM, et al. Effects of quality improvement strategies for type 2 diabetes on glycemic control: a meta-regression analysis. *JAMA.* 2006;296(4):427–40, Jul 26.
- Meigs JB, Cagliero E, Dubey A, et al. A controlled trial of web-based diabetes disease management: the MGH diabetes primary care improvement project. *Diabetes Care.* 2003;26(3):750–7, Mar.
- O'Connor PJ, Crain AL, Rush WA, Sperl-Hillen JM, Gutenkauf JJ, Duncan JE. Impact of an electronic medical record on diabetes quality of care. *Ann Fam Med.* 2005;3(4):300–6, Jul–Aug.
- Selby JV, Karter AJ, Ackerson LM, Ferrara A, Liu J. Developing a prediction rule from automated clinical databases to identify high-risk patients in a large population with diabetes. *Diabetes Care.* 2001;24(9):1547–55, Sep.
- Murray D. Design and Analysis of Group-Randomized Trials. New York: Oxford University Press; 1998.
- Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. London: Arnold; 2000.
- Eldridge SM, Ashby D, Feder GS, Rudnicka A, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials.* 2004;1(1):80–90, Feb.
- Simpson JM, Klar N, Donnor A. Accounting for cluster randomization: a review of primary prevention trials, 1990 through 1993. *Am J Public Health.* 1995;85(10):1378–83, Oct.
- Osheroff JA, Teich JM, Middleton BF, Steen EB, Wright A, Detmer DE. A Roadmap for National Action on Clinical Decision Support. *J Am Med Inform Assoc.* 2007;14:141–5.
- Eccles M, McColl E, Steen N, et al. Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. *BMJ.* 2002;325(7370):941, Oct 26.
- Campbell MK, Grimshaw JM, Elbourne DR. Intraclass correlation coefficients in cluster randomized trials: empirical insights into how should they be reported. *BMC Med Res Methodol.* 2004;4:9, Apr 28.
- Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intraclass correlation coefficient in cluster randomized trials: the case of implementation research. *Clin Trials.* 2005;2(2):99–107.
- Lancaster GA, Chellawamy H, Taylor S, Lyon D, Dorrwick C. Design of a clustered observational study to predict emergency admissions in the elderly: statistical reasoning in clinical practice. *J Eval Clin Pract.* 2007;13(2):169–78, Apr.
- Johnson DW, Craig W, Brant , Mitton C, Svenson L, Klassen TP. A cluster randomized controlled trial comparing three methods of disseminating practice guidelines for children with croup. *Implement Sci.* 2006;1:10.
- Freedman LS, Gail MH, Green SB, Corle DK. The efficiency of the matched-pairs design of the Community Intervention Trial for Smoking Cessation (COMMIT). *Control Clin Trials.* 1997;18(2):131–9, Apr.
- Glasgow RE, Nutting PA, King DK, Nelson CC, Cutter G, Gaglio B, et al. Randomized effectiveness trial of a computer-assisted intervention to improve diabetes care. *Diabetes Care.* 2005;28(1):33–9, Jan.
- Grosskurth H, Mosha F, Todd J, Mwijarubi E, Klokke A, Senkoro K, et al. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. *Lancet.* 1995;346(8974):530–6, Aug 26.
- Gulmezoglu AM, Villar J, Grimshaw J, Piaggio G, Lumbiganon P, Langer A. Cluster randomized trial of an active, multifaceted information dissemination intervention based on The WHO Reproductive health library to change obstetric practices: methods and design issues. *BMC Med Res Methodol.* 2004;4(1):2, Jan 15.
- Thompson SG, Pyke SD, Hardy RJ. The design and analysis of paired cluster randomized trials: an application of meta-analysis techniques. *Stat Med.* 1997;16(18):2063–79, Sep 30.
- American Diabetes Association. Standards of medical care for patients with diabetes mellitus. *Diabetes Care.* 2003;26(1):S33–50, Jan.
- Raab GM, Butcher I. Balance in cluster randomized trials. *Stat Med.* 2001;20(3):351–65, Feb 15.

24. **Raab GM, Butcher I.** Randomization inference for balanced cluster-randomized trials. *Clin Trials*. 2005;2(2):130–40.
25. **Cohen J.** *Statistical Power Analysis for the Behavioral Sciences*. Toronto: Academic Press, Inc.; 1977.
26. **Ahmed A, Husain A, Love TE, et al.** Heart failure, chronic diuretic use, and increase in mortality and hospitalization: an observational study using propensity score methods. *Eur Heart J*. 2006;27(12):1431–9, Jun.
27. **Normand ST, Landrum MB, Guadagnoli E, et al.** Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol*. 2001;54(4):387–98, Apr.
28. **Kerry SM, Bland JM.** The intraclass correlation coefficient in cluster randomisation. *BMJ*. 1998;316(7142):1455, May 9.
29. **Feng Z, Diehr P, Peterson A, McLerran D.** Selected statistical issues in group randomized trials. *Annu Rev Public Health*. 2001;22:167–87.
30. **Kerry SM, Bland JM.** Sample size in cluster randomisation. *BMJ*. 1998;316(7130):549, Feb 14.
31. **Kerry SM, Bland JM.** Trials which randomize practices II: sample size. *Fam Pract*. 1998;15(1):84–7, Feb.
32. **Campbell MK, Elbourne D, Altman DG.** CONSORT statement: extension to cluster randomised trials. *BMJ*. 2004;328(7441):702–8, Mar 20.
33. **Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M.** Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract*. 2000;17(2):192–6, Apr.
34. **Campbell MK, Mollison J, Grimshaw JM.** Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Stat Med*. 2001;20(3):391–9, Feb 15.
35. **Cebul RD, Dawson NV, Love TE.** Cluster Randomized Trials in Health Care Research. In: Machin D, Day S, Green S (eds). *The Textbook of Clinical Trials 2nd ed*. New York: Wiley; 2006.
36. StataCorp. *Intercooled Stata/SE 9.2 for Windows*. College Station, TX: StataCorp; 2007.
37. **Puffer S, Torgerson D, Watson J.** Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ*. 2003;327(7418):785–9, Oct 4.
38. **Eccles M, Hawthorne G, Whitty P, et al.** A randomised controlled trial of a patient based Diabetes Recall and Management System: the DREAM trial: a study protocol. *BMC Health Serv Res*. 2002;2(1):5, Mar 21.
39. **Brown AF, Gregg EW, Stevens M, Karter AJ, Weinberger M, Safford MM, et al.** Race, ethnicity, socioeconomic position, and quality of care for adults with diabetes enrolled in managed care: the Translating Research Into Action for Diabetes (TRIAD) study. *Diabetes Care*. 2005;28(12):2864–70, Dec.
40. **Hsu J, Price M, Huang J, Brand , Fung V, Hui , et al.** Unintended consequences of caps on Medicare drug benefits. *N Engl J Med*. 2006;354(22):2349–59, Jun 1.
41. **Murray DM, Blitstein JL, Hannan PJ, Baker WL, Lytle LA.** Sizing a trial to alter the trajectory of health behaviours: methods, parameter estimates, and their application. *Stat Med*. 2007;26(11):2297–2316, May 20.