Comparative Effectiveness Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

A tool for assessing the feasibility of comparative effectiveness research

This article was published in the following Dove Press journal:
Comparative Effectiveness Research
29 January 2013
Number of times this article has been viewed

Alexander M Walker[1]
Amanda R Patrick[2]
Michael S Lauer[3]
Mark C Hornbrook[4]
Matthew G Marin[5]
Richard Platt[6]
Véronique L Roger[7]
Paul Stang[8]
Sebastian Schneeweiss[2]

[1]World Health Information Science Consultants, Newton, MA; [2]Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Boston, MA; [3]National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD; [4]The Center for Health Research, Kaiser Permanente Northwest, Portland, OR; [5]Department of Medicine, New Jersey Medical School, Newark, NJ; [6]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA; [7]Department of Health Sciences Research, Mayo Clinic, Rochester, MN; [8]Johnson and Johnson Pharmaceutical Research and Development, Titusville, NJ, USA

→ Video abstract

Point your smartphone at the QR code to the left. If you have a QR code reader the video abstract will appear. Or use: http://dvpr.es/WGSkxb

Correspondence: Alexander M Walker
275 Grove Street, Suite 2-400,
Newton, MA 02466, USA
Tel +1 617 663 5945
Email alec.walker@whiscon.com

Background: Comparative effectiveness research (CER) provides actionable information for health care decision-making. Randomized clinical trials cannot provide the patients, time horizons, or practice settings needed for all required CER. The need for comparative assessments and the infeasibility of conducting randomized clinical trials in all relevant areas is leading researchers and policy makers to non-randomized, retrospective CER. Such studies are possible when rich data exist on large populations receiving alternative therapies that are used as-if interchangeably in clinical practice. This setting we call "empirical equipoise."

Objectives: This study sought to provide a method for the systematic identification of settings it in which it is empirical equipoise that offers promised non-randomized CER.

Methods: We used a standardizing transformation of the propensity score called "preference" to assess pairs of common treatments for uncomplicated community-acquired pneumonia and new-onset heart failure in a population of low-income elderly people in Pennsylvania, for whom we had access to de-identified insurance records. Treatment pairs were considered suitable for CER if at least half of the dispensings of each treatment-pair member fell within a preference range of 30% to 70%.

Results: Among 3889 community-acquired pneumonia patients, insurance claims histories were sufficiently similar in seven drug pairs to suggest that observational CER might be effective. Relapse appears to have been less common in levofloxacin recipients than in similar patients given other products. In 6035 heart failure patients, metoprolol, carvedilol, and atenolol were employed in patients with similar claims histories, and thus might be suitable for observational CER. The long-acting succinate formulation of metoprolol had lower failure rates in head-to-head comparisons with all other beta-blockers. Both findings are candidates for further investigation. Confounding by unmeasured factors operating in the same manner as the measured covariates would not have produced the apparent superiority of levofloxacin, which was given to people in poorer respiratory health. The baseline covariate distributions of persons starting beta-blockers suggest only that carvedilol recipients were healthier than others.

Conclusion: A straightforward algorithm can identify empirical equipoise, in which prescribers as a group seem evenly divided on the merits of alternative therapies. This is the setting in which CER may be most necessary and is likely to be most accurate. The imbalances identified by propensity models can identify situations in which the results of screening analyses may be biased in the direction of the observed effect.

Keywords: equipoise, observational CER, methodology, community-acquired pneumonia, heart failure

Background

The Institute of Medicine recently used expert discussion and consensus to identify priorities for comparative effectiveness research (CER).[1] These were areas in which

treatments in common use might well have very different outcomes and for which there was little direct information. Perusal of these priority areas suggests that many of these will require a substantial and sophisticated research effort.

We present here an empirical method for identifying possible areas for CER needs. This method might be called the "low-hanging fruit algorithm" because it involves systematically looking for treatment alternatives for which the physician community appears to make very little distinction between which patients receive one treatment and which patients receive the other. This results in largely comparable treatment populations, which in turn can be compared for the occurrence of outcomes of all kinds, with only modest requirements for covariate control.

For example, clinical indifference might arise when: there is a lack of evidence to differentiate two products, there is positive evidence of equivalence, there is an expectation of equivalence because of common pharmacology or membership in a therapeutic class, guidelines give equal weight to therapeutic alternatives, or the net effect of promotion and accessible scientific data has been to leave prescribers without a clear preferred treatment option. The manifestation of clinical indifference in patient data will be that no patient medical or demographic characteristics distinguish recipients of one therapy from those of another. We would like to use the term "empirical equipoise" to identify a setting of strong, observed similarity in the kinds of patients receiving two regimens.

"Equipoise" refers to a balance of opinion in the treating community about what really might be the best treatment for a given class of patients. The existence of equipoise can make a randomized trial ethically justifiable even when no single researcher or clinician is personally in doubt about which might be the best course of treatment; that is, a treating professional can think: "Seeing that others who are equally informed disagree with me, I recognize that we as a treatment community are in a state of equipoise." Empirical equipoise differs from true equipoise in that the balance of prescribers' actions is taken as the measure of preference rather than their opinions.

Under empirical equipoise, individual physicians may have preferences for treatments for particular patients, but the preferences tend to balance out across providers. Not knowing the opinions of prescribers, we cannot say whether there is true equipoise, but we can observe that the prescriber community appears to be acting as if there were equipoise. Empirical equipoise is the practice condition in which comparative observational studies can be pursued with a diminished concern for the "confounding by indication" that plagues non-randomized studies of treatment effect.

The very large bodies of information on medical care and outcomes in governmental and private insurance schemes offer an avenue for prioritizing topics for CER. We can, at least for those settings, identify alternative therapies for which there is empirical equipoise and for which observational, retrospective CER might be most feasible.

Methods

A prioritization tool

We propose the following algorithm.

1. Identify an environment with longitudinal health care data for a large population in which CER may be relevant. The data source must credibly capture the critical variables to define the date and nature of the treatment (eg, drug name, days supplied, date prescription filled, dosage, quantity dispensed), likely indications, and relevant outcomes (these might include diagnosis codes, procedure codes, diagnostic test codes, hospital discharge diagnoses).
2. Select treatment indications that have a clear onset that can be dated using the available data, for which there are a variety of treatment options that might be compared, and which are also clearly captured in the available data.
3. Taking all indications pair-wise, perform a regression in which treatment choice with one member of the pair (A vs B) is the dependent variable and the predictors are an exhaustive list of patient characteristics, potential predictors of the outcome, and patterns of care identified before the treatment decision.
4. From the predicted values of the regression, create a preference score ranging from 0 to 1 that has the properties that:
   a. patients with preference scores of 0 or 1 receive Treatment A either never or always, respectively
   b. intermediate values of the preference score reflect the proportion of patients who would be expected to receive Treatment A rather than Treatment B, under the circumstance that Treatment A and Treatment B had equal market share
   c. (as a consequence of b) patients whose preference score is 0.5 are likely to receive Treatment A or B exactly in proportion to the market shares of Treatments A and B.
5. Accept drug pairs as emerging from empirical equipoise if at least half of the dispensings of each of the drugs are to patients with a preference score of between 0.3 and 0.7.

Data source for case studies

We selected records of patients from the Pennsylvania Pharmacy Assistance Contract for the Elderly (PACE) linked to Medicare Parts A and B claims from 2000 through 2005. To be eligible for PACE at the time, a person's annual income had to be <US$13,000 if single and <US$16,200 if married. The program reimburses the cost of all prescriptions with a US$6 co-payment. To our knowledge, there were no formulary restrictions for antibiotics or beta-blockers in the program for this period.

PACE's computerized records include prescription drug name, dosage, quantity dispensed, days supplied, date dispensed, and a code indicating the prescribing physician. Outpatient and inpatient diagnoses, procedure codes, and dates of all inpatient and outpatient services obtained from Medicare claims data were linked to pharmacy dispensing data and to Pennsylvania vital-statistics files. To protect the privacy of subjects and their physicians, all personal identifiers had already been transformed into de-indentified study numbers in the study files to which we had access. We obtained ethics approval from the Brigham and Women's Hospital Institutional Review Board and data use agreements from PACE and the Center for Medicare and Medicaid Services to use their data.

Preference score

To quantify the degree of overlap in physician choice between two drugs, we employed a propensity score whose components could be identified by an algorithm. There were two motivations for an algorithmic rather than expert-derived score; we wished to: (1) mimic a screening project in which a data source could be reviewed for a large number of possible treatment-pair candidates for CER and (2) take advantage of the rich predictor data and a large number of observations.

The propensity scores for persons receiving Treatment A versus those receiving Treatment B are the fitted values of the logistic regression derived in comparing persons receiving Treatment A to those receiving B. The components of the propensity score for the screening exercise described here were indicator variables for age in decades on index date (see later in this paper for case-specific definitions of the index date); sex; three-digit *International Classification of Diseases, 9th Revision – Clinical Modification* (ICD-9)[2] codes appearing as a principal hospital diagnosis in the periods 1–7 and 8–270 days before the index date (retained if found in at least 2% of the population); three-digit ICD-9 codes appearing in association with a physician outpatient visit in the periods 1–7 and 8–270 days before the index date (retained if found in at least 2% of the population); and the following general indicators of health in the 270 days before the index date: number of drugs used (at distinct generic entity level),[3] number of physician visits, indicator (Y/N) for hospitalization, indicator (Y/N) for nursing home admission, and Charlson comorbidity score.[4,5]

We derived a transformation of the propensity score that would be more interpretable as a measure of the preference for one drug or another associated with patient demographic and health characteristics. This preference score was obtained by subtracting the natural logarithm of Treatment A prevalence divided by Treatment B prevalence from the logit of the propensity score, and taking the anti-logit (expit) of the result. In the resulting equation (Equation 1), in the universe of persons receiving either Treatment A or B, F and S are the preference score and propensity score for receiving Treatment A, respectively, and P is the fraction of persons receiving Treatment A:

$$\ln\left(\frac{F}{1-F}\right)=\ln\left(\frac{S}{1-S}\right)-\ln\left(\frac{P}{1-P}\right) \quad (1)$$

To confirm this works in an extreme case, imagine that no patient predictors differentiate use of Treatment A from use of Treatment B, and that the ratio of prevalence of Treatment A to Treatment B is 1/9 (ie, Treatment A has a 10% share of the treatments that are either A or B). All the coefficients of the propensity score would be zero and all subjects would have a propensity score of 0.10. The intercept of the regression equation would be ln[0.1/(1–0.1)] = –2.20. The procedure for deriving a preference score would subtract ln(1/9) = –2.20 from the logit of the propensity score of each subject, making each logit preference score zero. A zero logit preference corresponds to a 50% preference, since logit(0.5) = 0. Everybody would correctly get a preference score of 0.50 because nobody would possess a characteristic that made treatment with A more or less likely than use of Treatment A overall.

Case study 1: community-acquired pneumonia (CAP)

The Infectious Disease Society of North America guidelines identify persons without chronic disease predisposing to drug-resistant *Streptococcus pneumoniae* infection and without recent antibiotic therapy as a population for whom optimal treatment of CAP involves a relatively small number of options.[6] This screening exercise aims to identify all commonly used treatments in this setting and to compare them for preliminary evidence of disparities in outcome.

Methods

Study subjects

PACE recipients were included in this analysis if they had all three of the following findings in their insurance files:

- an outpatient physician visit bearing ICD-9 codes 482.9 (bacterial pneumonia unspecified), 485 (bronchopneumonia, organism unspecified), or 486 (pneumonia, organism unspecified) following at least 270 days of uninterrupted recorded data in PACE
- a radiologic exam of the chest up to 3 days before or 2 days after the day of diagnosis. These are marked in insurance files by CPT-4 codes 71010 (single view, frontal), 71015 (stereo, frontal), and 71021 (two views, frontal and lateral) as well as ICD-9 87.44 (routine chest X-ray, so described, including X-ray of chest not otherwise specified)
- a dispensing from 3 days before to 3 days after the day of diagnosis of single antibiotics indicated or used for CAP. We began with a broad list of all systemically absorbed antibiotics, but restricted analysis to those given to at least 5% of the CAP patients. The few individuals who received multiple antibiotics were not included in the analysis.

Individuals who qualified for selection on multiple occasions were retained only the first time they qualified for inclusion in the study. The date of initiation of antibiotic treatment was taken as Day 0. Persons meeting all of the just-outlined criteria were excluded from further study if they had:

- an insurance claim during the 270 days prior to Day 0 bearing ICD-9 code 042.x (human immunodeficiency virus infection)
- any dispensing in the preceding 3 months of insulin, or any of the antibiotics used to define treatment groups (as outlined previously)
- a discharge from an inpatient hospital stay with any diagnosis on Day −90 through Day 0
- an ICD-9 code on Day −30 through Day 0 of 480 (viral pneumonia); 482 (pneumonia due to *Klebsiella pneumoniae*, *Pseudomonas*, *Haemophilus influenzae*, *Staphylococcus*, other specified bacteria); 483 (pneumonia due to *Mycoplasma pneumonia*, *Chlamydia*, other specified organism); 484 (pneumonia in cytomegalic inclusion disease, whooping cough, anthrax, aspergillosis, other systemic mycoses, other infectious diseases classified elsewhere); 487 (influenza); 488 (influenza due to identified avian influenza virus); pneumonia in other conditions: 518.3 (allergic or eosinophilic), 507 (aspiration), 770.0 (congenital), 514 (passive), 390 (rheumatic), 997.31 (ventilator associated).

Treatment failure

We identified treatment failure as the earliest (if any) occurrence within 30 days after Day 0 of: hospitalization with either a principal discharge diagnosis of ICD-9 481 (pneumococcal pneumonia), 482.9 (bacterial pneumonia unspecified), 485 (bronchopneumonia, organism unspecified), 486 (pneumonia, organism unspecified, 510 (empyema), or 513.0 (abscess of lung) or a new dispensing of any antibiotic. Note that the hospitalization codes for treatment failure do not include 480 (viral pneumonia), 487 (influenza), or 488 (avian influenza).

Results

There were 30,291 cases of apparent CAP. From these, to identify uncomplicated cases as explained in the earlier section "Study subjects," we serially excluded: 989 patients with less than 270 days of baseline information, 13 with human immunodeficiency virus, 15,447 with recent antibiotic use, and 8611 with codes indicating a viral pneumonia or pneumonia due to an agent other than *S. pneumoniae*. This left 5231 that met the screening criteria for uncomplicated CAP. The antibiotics used by at least 5% of the uncomplicated cases were: azithromycin (28%), levofloxacin (27%), clarithromycin (7%), moxifloxacin (6%), and amoxicillin (5%). Agents not included, but with prevalence of use greater than 1% were: amoxicillin/clavulanate (4%), gatifloxacin (3%), ciprofloxacin (3%), cefuroxime (3%), doxycycline (3%), cephalexin (2%), sulfamethoxazole/trimethoprim (2%), and erythromycin (1%).

Table 1 shows the prevalence of demographic characteristics and the 25 most commonly noted baseline diagnoses by drug. Consistent with expectations for PACE recipients with pneumonia, this was an elderly population (mean age 82 years) of mostly women (80%). Despite the substantial filter for good health imposed by removing individuals with hospitalizations in the preceding 90 days, the population had high prevalence of previous respiratory and cardiac disease, as a well as a wide range of other chronic ailments. The antibiotic levofloxacin showed the highest prevalence of all common categories of respiratory conditions in the baseline 270 days. These included: respiratory and chest symptoms, pneumonia, and chronic obstructive pulmonary disease. Although receipt of isoniazid or pyrazinamide was not an exclusion criterion, none of the patients counted in Table 1 had received either of these drugs in the preceding 270 days.

Table 2 shows all the possible treatment comparisons, sorted by the proportion with preference values in the range

Table 1 Demographics, health care utilization, and outpatient diagnoses present in at least 10% of patients in the 270 days before initiation of treatment for community-acquired pneumonia

Patient demographics		Amoxicillin	Azithromycin	Clarithromycin	Levofloxacin	Moxifloxacin
Patients, n		269	1468	369	1407	320
Age in years, mean		82	82	81	82	81
Female, %		81	82	79	78	76
Prior nursing home admission, %		5	4	3	8	4
Prior hospitalization, %		19	19	15	22	17
Distinct drugs dispensed, mean		8	9	8	9	9
Physician visits, mean		6	7	6	7	7
Comorbidity score, mean		2	2	2	2	2
Outpatient diagnoses, %						
ICD-9	*1–7 days before first dispensing*					
486	Pneumonia, organism unspecified	6	16	13	23	13
786	Symptoms involving respiratory system chest	6	16	12	20	11
	8–270 days before first dispensing					
401	Essential hypertension	53	55	53	51	54
110	Dermatophytosis	34	28	22	29	27
272	Disorders of lipid metabolism	22	28	26	26	33
414	Other chronic ischemic heart disease	27	28	19	27	26
786	Symptoms of respiratory system and chest	24	25	22	29	22
V04	Need for prophylactic vaccination	24	24	27	25	24
780	General symptoms	18	22	18	21	21
715	Osteoarthrosis	25	20	18	22	21
729	Other disorders of soft tissues	21	20	16	21	22
250	Diabetes mellitus	22	20	20	19	23
427	Cardiac dysrhythmias	19	20	18	20	16
366	Cataract	19	18	18	20	14
496	Chronic airway obstruction NEC	12	18	15	20	20
443	Other peripheral vascular disease	19	18	13	19	16
719	Other and unspecified disorders of joint	17	17	12	19	19
428	Heart failure	19	17	14	16	15
244	Acquired hypothyroidism	10	15	9	15	13
733	Other disorders of bone and cartilage	12	13	9	14	19
362	Other retinal disorder	15	13	10	13	14
724	Other and unspecified disorders of back	14	13	13	13	10
365	Glaucoma	9	12	8	11	11
V58	Encounter for unspecified care	10	10	9	11	12
285	Other and unspecified anemias	7	11	9	11	9

Note: Data from Pennsylvania Pharmacy Assistance Contract for the Elderly (PACE), 2000–2005.
Abbreviations: ICD-9, International Classification of Diseases, 9th Revision – Clinical Modification[10]; NEC, not elsewhere classified.

of 0.3 to 0.7. The high proportions in the specified range for the first pair listed indicate that azithromycin and levofloxacin were given to the most similar groups of patients, while those in the last pair (amoxicillin and moxifloxacin) were given to the least similar groups. Figure 1 shows smoothed preference score distributions for a similar pair of treatment alternatives (azithromycin and levofloxacin) and a dissimilar pair (clarithromycin and moxifloxacin).

Table 3 presents the risk of presumed treatment failure between pairs of drugs with substantially similar usage, defined as having at least half the dispensings of both agents administered to patients who themselves were within a preference range of 0.3 to 0.7. Crude and preference score-adjusted effect estimates are very similar, and point to a lower risk of treatment failure in users of levofloxacin compared with each of the other drugs.

Case study 2: heart failure

The American College of Cardiology, the American Heart Association, and the International Society for Heart and Lung Transplantation have provided guidelines calling for use of beta-blockers in patients with heart failure and depressed left-ventricular systolic function.[6] The guidelines identify three beta-blockers that have been shown to improve survival: bisoprolol, sustained-release metoprolol, and carvedilol. Other beta-blockers have either not been tested or failed to improve survival in randomized trials. There has been only one large-scale trial comparing different beta-blockers,

Table 2 Numbers of patients and preference overlap for antibiotic pairs taken by at least 5% of patients with community-acquired pneumonia

Antibiotic pair	Patients, n	0.3 ≤ preference ≤ 0.7	
		%	N
Azithromycin	1468	85	1254
Levofloxacin	1407	82	1159
Azithromycin	1468	61	889
Clarithromycin	369	66	244
Azithromycin	1468	59	867
Moxifloxacin	320	56	178
Clarithromycin	369	60	222
Levofloxacin	1407	55	779
Amoxicillin	269	60	162
Levofloxacin	1407	57	799
Amoxicillin	269	55	148
Azithromycin	1468	50	740
Levofloxacin	1407	53	752
Moxifloxacin	320	60	193
Clarithromycin	369	39	145
Moxifloxacin	320	42	133
Amoxicillin	269	43	116
Clarithromycin	369	43	159
Amoxicillin	269	39	104
Moxifloxacin	320	42	113

Note: Data from the Pennsylvania Pharmacy Assistance Contract for the Elderly program, 2000–2005.[10]

but it failed to use maximal doses for both agents.[7] Recent observational studies suggest no significant difference between sustained-release metoprolol and carvedilol or between "evidenced-based beta-blockers" (carvedilol, metoprolol succinate, and bisoprolol fumarate) and others among persons with established heart failure and a recent hospitalization.[8,9]

This screening exercise aimed to identify all commonly used beta-blockers in the setting of apparently new-onset heart failure and to compare them for preliminary evidence of disparities in outcome, doing this in a way that would be readily replicable in other populations.

Methods

Study subjects

PACE recipients were included in this analysis if they had all three of the following findings in their insurance files within a 28-day period:

- a dispensing of any of the following orally administered beta-blockers: acebutolol, atenolol, metoprolol tartrate, metoprolol succinate, carvedilol, bisoprolol, propranolol, sotalol, labetalol, pindolol, nadolol, timolol, or nebivolol
- an outpatient physician visit or principal hospital discharge diagnosis bearing the first instance in the patient record of ICD-9 code for heart failure (428) other than pure diastolic failure (428.3), following at least 365 days of uninterrupted recorded data in PACE
- evidence that ejection fraction was measured, as attested by the performance of echocardiography, gated single photon emission computed tomography myocardial perfusion imaging, multi-gated acquisition scan, contrast left ventriculography, cardiac magnetic resonance imaging, or fast scan cardiac computed tomography.

The date of the last occurring of these three events was taken as Day 0. Persons previously hospitalized and discharged with a diagnosis of heart failure were excluded (heart failure diagnoses were ICD-9 codes 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, or 428.9.)

Treatment failure

We identified treatment failure as either hospitalization by Day 60 with a principal discharge diagnosis of heart failure or death from any cause.

Results

There were 5247 cases of new-onset heart failure. The beta-blockers that comprised at least 5% of the initial prescriptions

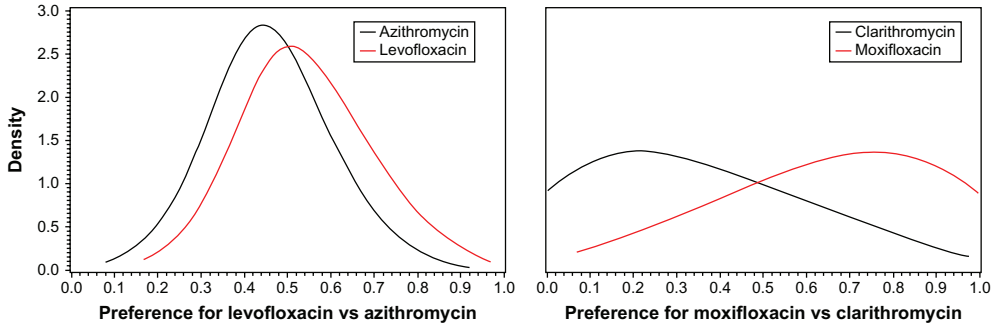

Figure 1 Preference score distributions.
Notes: Preference distributions for a pair of antibiotics given to very similar patients (left) and for a pair given to substantially different patient populations. (See Table 1 for salient differences).

Table 3 Relative risk of presumed treatment failure in community-acquired pneumonia elders

Antibiotic pair	Patients, n	Failures, n	%	Relative risk, crude	PS-adjusted odds ratio
Levofloxacin	1159	337	29	0.72	0.62
Azithromycin	1254	508	41		(0.52–0.74)
Clarithromycin	244	93	38	0.99	0.96
Azithromycin	889	342	38		(0.71–1.30)
Azithromycin	740	304	41	0.67	0.42
Amoxicillin	148	91	61		(0.29–0.61)
Moxifloxacin	178	60	34	0.83	0.70
Azithromycin	867	352	41		(0.49–0.99)
Levofloxacin	779	222	28	0.71	0.59
Clarithromycin	222	89	40		(0.43–0.81)
Levofloxacin	799	258	32	0.54	0.32
Amoxicillin	162	96	59		(0.22–0.45)
Levofloxacin	752	225	30	0.95	0.95
Moxifloxacin	193	61	32		(0.67–1.34)

Note: Data from the Pennsylvania Pharmacy Assistance Contract for the Elderly program, 2000–2005.[10]
Abbreviation: PS, preference score.

were metoprolol succinate (the extended-release formulation of metoprolol in the USA), metoprolol tartrate, carvedilol, and atenolol. Carvedilol recipients included somewhat fewer women and fewer prior nursing home residents than patients who received the other beta-blockers. Further, fewer carvedilol recipients had had care associated with any of the common diagnoses shown Table 4, which includes all diagnoses seen in at least 10% of the population. Table 5, which shows that all the drug pairs met our criterion for comparability, nonetheless confirms the impression that carvedilol usage differed modestly from that of the other agents: all the most distinct drug pairs (those with a lower preference overlap) involved carvedilol.

Table 6 presents the risk of presumed treatment failure between patients in the overlapping range of preferences between the two products. Pairs of beta-blockers are ordered in Table 6 as in Table 5. Metoprolol succinate, the extended-release formulation of metoprolol, appears to carry a lower risk of treatment failure than the other three products, and metoprolol tartrate a lower risk than carvedilol.

Discussion

Empirical equipoise is at the heart of this tool for identifying feasible areas for CER. We propose a systematic method for finding pairs of treatments in which identified patient characteristics do not seem to play an important role in determining which patient receives which therapy. If patient characteristics are not determinative, it follows that selection has been driven by administrative factors (such as varying formularies and copayments or institutional treatment guidelines) and physician preferences (whether these are based on empirical evidence or sales persuasion is irrelevant for the present purpose). To the extent that non-patient determinants of therapy are also not determinative of outcome, they can play a role analogous to randomization in forming comparative groups. Randomization and patient-unrelated treatment assignment share the key characteristic of not resulting in any systematic expectation of greater success of one treatment over the other.

An attractive aspect of the comparative analysis of patients with overlapping preference scores is that the balance in the compared populations is a baseline characteristic. If the compared agents have different patterns of effects (not just treatment success but side effects, for example, or even side effects associated with particular patient attributes) the baseline balance in patient attributes gives confidence that known factors do not confound the drug-to-drug comparison.

The suitability for CER of treatments whose assignment appears to be unrelated to patient attributes depends on the balance and nature of unmeasured predictors of treatment choice. Preference, like the propensity score on which it is based, incorporates only those patient attributes that are known in the data system. Factors such as patient preference, patient experience with other medications, and even the patient–physician relationship will affect the choice of medications, particularly of those that require careful surveillance. By definition, we do not directly know about unmeasured predictors, but the clinical setting described measured predictors can be strongly suggestive. In the example presented of CAP treatment in elders who had not been recently hospitalized, the higher prevalence of respiratory illness in the histories of patients who received levofloxacin, rather than azithromycin, suggests that there is a systematic tendency for physicians to prefer levofloxacin in the presence of such a history. The measured predictors pointed to the levofloxacin patients being, if anything, worse off at baseline, and it seems at least plausible that physicians might similarly have preferred levofloxacin to azithromycin in patients with other unmeasured baseline markers of poor prognosis. In the example of initial treatment for heart failure, the somewhat healthier status of the carvedilol recipients militates against residual confounding as the explanation for its apparent inferiority to both formulations of metoprolol. There is little in the summary patient characteristics to suggest a tendency for persons with less severe underlying disease to have received metoprolol succinate than other products, so again an explanation by confounding seems not

Table 4 Demographics, health care utilization, recent inpatient diagnoses and outpatient diagnoses present in at least 10% of patients in the 270 days before initiation of treatment for heart failure

Patient demographics		Atenolol (Tenormin®)	Carvedilol (Coreg®)	Metoprolol succinate (Toprol-XL®)	Metoprolol tartrate (Lopressor®)
Patients, n		619	907	1164	2303
Age in years, mean		82	81	81	82
Female, %		82	72	78	80
Prior nursing home admission, %		36	26	34	40
Prior hospitalization, %		4	3	3	4
Distinct drugs dispensed, mean		7	7	7	7
Physician visits, mean		4	4	4	4
Comorbidity score, mean		2	2	2	2
Inpatient diagnoses					
ICD-9	1–7 days before day 0				
410	Acute myocardial infarction	4	3	7	8
414	Other forms of chronic ischemic heart disease	3	0	2	3
427	Cardiac dysrhythmias	5	2	3	3
Outpatient diagnoses					
1–7 days before day 0					
786	Symptoms of respiratory system and chest	22	17	28	26
401	Essential hypertension	19	10	17	17
427	Cardiac dysrhythmias	15	10	15	14
414	Other forms of chronic ischemic heart disease	13	9	13	16
780	General symptoms	10	7	11	12
	8–270 days before first dispensing				
401	Essential hypertension	56	51	55	55
250	Diabetes mellitus	30	33	29	31
786	Symptoms of respiratory system and chest	28	26	29	28
272	Disorders of lipoid metabolism	28	27	30	27
110	Dermatophytosis	29	25	28	28
414	Other forms of chronic ischemic heart disease	23	27	24	26
715	Osteoarthrosis	25	21	24	24
780	General symptoms	23	20	23	22
V04	Need for prophylactic vaccination	22	20	19	21
729	Other disorders of soft tissues	19	20	18	20
719	Other and unspecified disorders of joint	22	17	18	19
443	Other peripheral vascular disease	18	17	20	19
427	Cardiac dysrhythmias	18	19	21	17
366	Cataract	20	15	20	19
362	Other retinal disorder	17	15	16	17
724	Other and unspecified disorders of back	14	10	15	14
244	Acquired hypothyroidism	13	12	12	12
365	Glaucoma	13	9	14	12
429	Ill-defined heart disease	12	11	13	12
285	Other and unspecified anemias	10	12	11	13
733	Other disorders of bone and cartilage	9	9	12	11
789	Other symptoms of abdomen and pelvis	12	10	9	12
496	Chronic airway obstruction NEC	9	12	9	12
782	Symptoms involving skin and integument	12	10	11	10
V58	Encounter for unspecified care	9	12	10	10

Note: Data from Pennsylvania Pharmacy Assistance Contract for the Elderly (PACE), 2000–2005.
Abbreviations: ICD-9, International Classification of Diseases, 9th Revision – Clinical Modification[10]; NEC, not elsewhere classified.

well supported. In both examples, the distinctions identified in the screening procedure deserve to be the object of more rigorous investigation.

Both examples considered here have to do with initial therapy for an acute or evolving condition. For many second-line therapies and for treatments whose implementation is not related to an identifiable clinical episode, the weight of unmeasured determinants of treatment initiation may be too great to derive presumptive equipoise from a similarity in measured characteristics.

Table 5 Pairs of beta-blockers for which >50% of patients on each drug have a preference score between 0.3 and 0.7

Beta-blocker pair	Total	0.3 ≤ preference ≤ 0.7	
	Patients, n	%	N
Metoprolol tartrate	2303	83%	1917
Atenolol	619	85%	527
Metoprolol succinate	1164	86%	998
Metoprolol tartrate	2303	86%	1983
Metoprolol succinate	1164	73%	850
Atenolol	619	71%	442
Metoprolol tartrate	2303	71%	1641
Carvedilol	907	75%	682
Metoprolol succinate	1164	67%	781
Carvedilol	907	73%	661
Carvedilol	907	63%	575
Atenolol	619	59%	367

Finding equipoise in a population means that it may be a suitable venue for unbiased research, but without contextual knowledge it may be difficult to distinguish two scenarios that have different implications for the need of additional CER: (1) equipoise may stem from a belief that treatments are interchangeable, which is widely held despite an absence of supporting evidence; (2) equipoise may exist because it is scientifically well established that treatments are indeed equally effective. Only the former scenario indicates a need for more CER.

Table 6 Relative risk of presumed treatment failure in new-onset heart failure in elders

Beta-blocker pair	Patients	Failures	%	Relative risk	PS-adjusted odds ratio
Metoprolol tartrate	917	310	16%	0.93	0.91
Atenolol	527	92	17%		(0.70–1.18)
Metoprolol succinate	998	138	14%	0.83	0.82
Metoprolol tartrate	1983	331	17%		(0.66–1.03)
Metoprolol succinate	850	109	13%	0.71	0.66
Atenolol	442	80	18%		(0.48–0.91)
Metoprolol tartrate	1641	266	16%	0.80	0.78
Carvedilol	682	138	20%		(0.62–0.99)
Metoprolol succinate	781	103	13%	0.68	0.61
Carvedilol	661	128	19%		(0.45–0.82)
Carvedilol	575	105	18%	0.87	0.81
Atenolol	367	77	21%		(0.57–1.13)

Note: Patients in each comparison are restricted to those preference scores in the range of 0.3 to 0.7 for that pair.

Data from the Pennsylvania Pharmacy Assistance Contract for the Elderly program, 2000–2005.[10]

Abbreviation: PS, preference score.

Participants in the PACE program are generally female, economically poor, very elderly, and live in Pennsylvania. Would the conclusions of our comparative analyses carry over to young men working in New Mexico? The answer does not lie within the PACE data, but is a matter of scientific generalization. Our speculation is that the individuals whose health events we have assessed here may be different in fundamental ways relevant to drug effectiveness from people who are younger, economically more wealthy, and living in other regions and with differently structured systems of medical care. Accordingly, observational CER needs to be undertaken in many different populations, and the results – particularly if they vary – interpreted with care. The goal remains to search out reproducible findings that correspond to generalizable medical facts.

This exercise contrasts with the humorous joke about the inebriated researcher looking for his car keys under a streetlight after a late faculty party. "Did you lose them here?" a colleague asks. "No," comes the reply, "but this is the only spot that I can see anything." To extend the metaphor, what we propose here is a way to map out locations of lights. That comparative research could be done well under a figurative streetlamp (that is, in largely comparable groups) does not mean that it should be done – the addressable issues under the light may be unimportant. However, if we prioritize research according to medical, financial, or societal goals, and as a result find that we would need to undertake non-randomized effectiveness research in the "darkness" of non-comparable groups, we might prefer to take along a flashlight instead: randomization.

The covariates that we included in the preference score calculation were ones that could be derived from insurance claims, but the technique could be applied in any setting in which there is extensive information. If one had an electronic medical record, perhaps supplemented by descriptors of patient lifestyle choices such as smoking, alcohol drinking, or exercise, much more information could have gone into the estimation of whether alternative treatments really were being delivered to substantially similar patients.

PACE, the venue for the examples we have explored, consists of a single social stratum of individuals of relatively homogenous age in a single geographic area, with health care claims data generated under a common reimbursement structure. Observational CER studies that include more heterogeneous data may encounter treatment technologies that vary according to location and other determinants of outcome, not all of them well represented in the claims histories or electronic health records. Depending on the source

and the time, the data may include fraudulent care, unjustified claims, defective products, and impaired physicians, and may miss addictions or other hidden comorbidities, all to different degrees and possibly differentially associated with the technologies to be compared. Our methods strive to infer equipoise from the comparability of group characteristics in the recipients of different therapy. Social stratification influences who seeks care when, how treatments are delivered, how patients adhere to those treatments, and whether outcomes are even recognized. Observational CER needs to be undertaken in many different populations, and the results – particularly if they vary – interpreted with care.

Acknowledgments

This report is the product of a working group of the Clinical Effectiveness Research Innovation Collaborative. The Collaborative is convened under the auspices of the Institute of Medicine's Roundtable on Value and Science-Driven Health Care to provide a venue for researchers working together to develop and apply innovative approaches to improving the pace and progress of clinical effectiveness research. The work was coordinated and overseen by LeighAnne Olsen, PhD.

Other members of the working group, who contributed to the development of this paper through discussion, review, and constructive suggestions, were Lawrence J Fine and Cato Laurencin. Michael Klompas and Deborah Yokoe reviewed interim versions of this report and provided helpful feedback.

Disclosure

The authors declare no conflicts of interest in this work.

References

1. Institute of Medicine Committee on Comparative Effectiveness Research Prioritization. *Initial National Priorities for Comparative Effectiveness Research*. Washington DC: National Academies Press; 2009. Available from: http://www.nap.edu/catalog.php?record_id=12648. Accessed February 19, 2010.
2. World Health Organization. *International Classification of Diseases, 9th Revision – Clinical Modification*. Geneva: World Health Organization; 1998.
3. Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. 2001;154(9): 854–864.
4. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol*. 1992;45(6):613–619.
5. Mandell LA, Wunderink RG, Anzueto A, et al; Infectious Diseases Society of America; American Thoracic Society. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Inf Dis*. 2007;44 Suppl 2:S27–S72.
6. Jessup M, Abraham WT, Chin MH, et al; American College of Cardiology Foundation; American Heart Association. 2009 Focused update incorporated into the ACC/AHA 2005 Guidelines for the Diagnosis and Management of Heart Failure in Adults A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines Developed in Collaboration With the International Society for Heart and Lung Transplantation. *J Am Coll Cardiol*. 2009;53(15):e1–e90.
7. Poole-Wilson PA, Swedberg K, Cleland JG, et al; COMET investigators. Comparison of carvedilol and metoprolol on clinical outcomes in patients with chronic heart failure in the Carvedilol Or Metoprolol European Trial (COMET): randomised controlled trial. *Lancet*. 2003;362(9377): 7–13.
8. Go AS, Yang J, Gurwitz JH, Hsu J, Lane K, Platt R. Comparative effectiveness of different beta-adrenergic antagonists on mortality among adults with heart failure in clinical practice. *Arch Intern Med*. 2008; 168(22):2415–2421.
9. Kramer JM, Curtis LH, Dupree CS, et al. Comparative effectiveness of beta-blockers in elderly patients with heart failure. *Arch Intern Med*. 2008;168(22):2422–2428.
10. International Classification of Diseases Clinical Modification (ICD-9-CM) October 2011 (CD ROM) US Department of Health and Human Services (DHHS) Publication PHS 11-1260. NCHS CD ROM 2011, No. 1. Available from http://bookstore.gpo.gov/actions/GetPublication.do?stocknumber=017-022-01616-8. Accessed January 8, 2013.