Why Does Primary Angioplasty Not Work in Registries? Quantifying the Susceptibility of Real-World Comparative Effectiveness Data to Allocation Bias

Sayan Sen, MRCP; Justin E. Davies, MRCP; Iqbal S. Malik, FRCP; Rodney A. Foale, FRCP; Ghada W. Mikhail, FRCP; Nearchos Hadjiloizou, MRCP; Alun Hughes, MRCP; Jamil Mayet, FRCP; Darrel P. Francis, FRCP

Background—Meta-analysis of registries (comparative effectiveness research) shows that primary angioplasty and fibrinolysis have equivalent real-world survival. Yet, randomized, controlled trials consistently find primary angioplasty superior. Can unequal allocation of higher-risk patients in registries have masked primary angioplasty benefit?

Methods and Results—First, we constructed a model to demonstrate the potential effect of allocation bias. We then analyzed published registries (55,022 patients) for allocation of higher-risk patients (Killip class ≥1) to determine whether the choice of reperfusion therapy was affected by the risk level of the patient. Meta-regression was used to examine the relationship between differences in allocation of high-risk patient to primary angioplasty or fibrinolysis and mortality. Initial modeling suggested that registry outcomes are sensitive to allocation bias of high-risk patients. Across the registries, the therapy receiving excess high-risk patients had worse mortality. Unequal distribution of high-risk status accounted for most of the between-registry variance (adjusted $R^2_{meta}$ = 83.1%). Accounting for differential allocation of higher-risk patients, primary angioplasty gave 22% lower mortality (odds ratio, 0.78; 95% confidence interval, 0.64–0.97; $P=0.029$). We derive a formula, called the number needed to abolish, highlighting situations in which comparative effectiveness studies are particularly vulnerable to this bias.

Conclusions—In ST-segment elevation myocardial infarction, clinicians’ preference for management of a few high-risk patients can shift mortality substantially. Comparative effectiveness research in any disease is vulnerable to this, especially diseases with an immediately identifiable high-risk subgroup that clinicians prefer to allocate to 1 therapy. For this reason, preliminary indications from registry-based comparative effectiveness research should be definitively tested by randomized, controlled trials.  (Circ Cardiovasc Qual Outcomes. 2012;5:00–00.)

Key Words: STEMI ■ primary angioplasty ■ fibrinolysis ■ allocation bias

Randomized, controlled trials (RCTs) and meta-analyses of these trials, on the treatment of ST-segment elevation myocardial infarction (STEMI) have consistently shown that patients randomly allocated to primary angioplasty (PPCI) have lower mortality than those randomly allocated to fibrinolysis.\(^1\) It is therefore increasingly adopted as a standard treatment for STEMI, often requiring reorganization of regional emergency infrastructure to allow PPCI delivery around the clock.\(^1\)

Real-world observational experience, however, has not appeared so favorable for PPCI. These studies are often larger and describe a more comprehensive spectrum of STEMI patients. Many show no significant reduction in mortality in the PPCI group compared with the fibrinolysis group.\(^5\)–\(^12\) A meta-analysis of such registries has revealed no significant difference in long-term mortality between patients allocated to PPCI and those allocated to fibrinolysis.\(^13\) The greater size, universal inclusivity, and wider geographical spread of real-world data form a significant challenge to the RCT-based assumption that PPCI does indeed give genuine long-term benefits.\(^14\)

One possible explanation for the paradox is that selective inclusion in RCTs may have drawn from a restricted pool of patients who benefit from PPCI and eliminated a large group who do not benefit or even suffer harm. If true, the international push to provide PPCI facilities to all is misguided. An alternative explanation is that, in registries, patients are allocated to therapy with some regard to their clinical status. For example, in some environments, patients with a higher risk of death (eg, cardiogenic shock) may receive preferential allocation to PPCI rather than fibrinolysis. In other environments, interventional cardiologists may be cautious about causing procedural harm or requiring intersite transport in higher-risk patients who may thereby preferentially be allocated to fibrinolysis rather than PPCI.

Received September 19, 2012; accepted September 27, 2012.
From the Imperial College Healthcare National Health Service Trust, St. Mary’s Hospital, Imperial College London, London, UK.

The online-only Data Supplement is available at http://circoutcomes.ahajournals.org/lookup/suppl/doi:10.1161/CIRCOUTCOMES.112.966853/-/DC1.
Correspondence to Sayan Sen, International Center of Circulatory Health, National Heart and Lung Institute, 59–61 N Wharf Rd London, W2 1LA, UK. E-mail sayan.sen@imperial.ac.uk

© 2012 American Heart Association, Inc.

Circ Cardiovasc Qual Outcomes is available at http://circoutcomes.ahajournals.org
DOI: 10.1161/CIRCOUTCOMES.112.966853
Although this is a recognized limitation of registries, the consequences of such allocation bias on registry results in this field have never been quantified.

We scrutinized the distribution of higher-risk patients (Killip class ≥1) across several STEMI registries to quantify whether they are preferentially allocated to a particular reperfusion therapy and to evaluate whether any unequal distribution of these patients could be enough to explain the conflict between RCTs and registries.

First, we constructed a model using values from prior studies to quantify, simplistically, how easily unequal allocation of higher-risk patients to an intrinsically better therapeutic arm can appear to abolish its benefit.

Second, we analyzed the 11 registries incorporated into the recent meta-analysis documenting nonsuperiority of PPCI outcomes. We examined how higher-risk patients were allocated between PPCI and fibrinolysis to see whether studies allocated higher-risk patients equally between therapies and whether any inequality was associated with any differences in mortality between therapies. We performed a weighted, random-effects meta-regression of the registries, taking into account higher-risk patient allocation, to estimate the odds ratios of mortality (akin to an informal RCT in which higher-risk patients were allocated equally to both therapies). The analysis was then repeated for other recorded covariates likely to affect outcomes (female sex, hypertension, diabetes mellitus, and previous stroke) to determine whether these other variables could also contribute to the paradox.

Finally, we developed a general formula for the “number needed to abolish (NNA)” and discuss its implications.

WHAT IS KNOWN

• There is conflict between randomized, controlled trials demonstrating that primary angioplasty is superior to fibrinolysis in ST-segment elevation myocardial infarction, and observational, registry-based comparative effectiveness research contradicting this.

• Understanding the basis of this conflict could help determine how future comparisons between therapies should be conducted.

WHAT THE STUDY ADDS

• The conflict between randomized, controlled trials and observational research may be explained by preferential allocation of higher risk patients by clinicians to primary angioplasty.

• We derived a formula for resistance to allocation bias of observational results, the ‘Number needed to Abolish’, and demonstrate that it is small for myocardial infarction.

• The results of this study suggest that observational comparative efficacy research is especially vulnerable to incorrect conclusions when clinicians (a) can readily identify a high-risk subset, and (b) preferentially allocate them to one therapy rather than the other.

Methods

Simple Calculation of Number Needing to Be Diverted to Abolish Apparent Benefits of a Better Therapy

Initially, we prepared a model of an observational study of 200 STEMI patients in which higher-risk patients could be preferentially allocated to 1 arm or the other, to various extents. In this way, we tested the effect of very simple forms of unequal allocation of higher-risk patients on the relative mortality of the 2 arms.

The notional cohort of 200 patients was distributed 100 to PPCI and 100 to fibrinolysis. We set the prevalence of higher-risk characteristics (Killip class ≥1) to be 25% on the basis of the reported prevalence in STEMI cohorts. We considered the remaining 75% to be low risk. We then varied the distribution of higher-risk patients between arms. For example, the PPCI arm may receive 26 higher-risk patients plus 74 low-risk patients, in which case the fibrinolysis arm would receive the remaining 24 higher-risk patients plus 76 low-risk patients. We varied how the higher-risk patients were allocated between arms to test how easily the odds ratio can be influenced by biased allocation of this patient subgroup. Model parameters were based on published data (Table 1).

General Formula for How Many Patients Need to be Diverted to a Better Arm for Its Advantage to Disappear (NNA)

In the online-only Data Supplement we derive a general formula for the number of higher-risk patients per 100 total patients who need to be diverted from an intrinsically worse (control) arm to an intrinsically better (therapy) arm for the benefits of the better arm to disappear (NNA). This number turns out to be independent of the size of the study and mortality. What matters is:

• the difference in efficacy between arms, expressed as a relative risk reduction (RRR);

• the difference in risk between higher-risk and lower-risk patients, expressed as a relative risk increase (RRRI);

• the proportion of patients who are higher risk, f (bias).

The NNA is given by the following formula:

\[
NNA = \frac{RRR}{2 - RRR} \left[ \frac{1}{RRRI_{high}} + f \right]^{high}
\]

In real life, the relative risk between the higher-risk and other patients, as well as other parameters, may vary between different geographies and over the decades. We therefore also performed a sensitivity analysis by varying the model parameters (Table 2) to see how allocation bias in different settings would affect the relative mortality between therapies.

Analysis of Real-world Registries

We reanalyzed the set of 11 registries included in the thorough meta-analysis by Huynh et al., which found PPCI and fibrinolysis to have equivalent long-term survival. We scrutinized the distribution of higher-risk patients in the original studies. We defined higher risk as Killip class >1. If Killip class was not reported, the study’s own classification of higher risk was used. Only registries with sufficient baseline patient characteristic data to allow identification of higher-risk patients were included in our analysis. The difference in the proportions of higher-risk patients in the PPCI arm to the fibrinolysis arm was then calculated and plotted against the log odds ratio of mortality between therapies at 1 year. We also examined the distribution of other commonly reported covariates such as hypertension, stroke, diabetes mellitus, and female sex in a similar manner. A random-effects meta-regression using a restrictive maximum likelihood estimate of between-study variance with Knapp-Hartung modification was used to define the relationship between the ratio of higher-risk patients between therapies in each registry and mortality. An adjusted
Table 1. Simulated Model of 200 STEMI Patients Demonstrating How Allocation of Higher Risk Patients Between PPCI and Fibrinolysis Influences the Odds Ratio of Mortality

<table>
<thead>
<tr>
<th>Distribution of High-Risk Patients</th>
<th>Allocation</th>
<th>Expected Deaths</th>
<th>Allocation</th>
<th>Expected Deaths</th>
<th>Proportion of High-risk Cohort Misallocated</th>
<th>Difference in Prevalence of High Risk</th>
<th>Odds Ratio (&lt;1 PPCI Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of excess high-risk patients in PPCI</td>
<td>3</td>
<td>22</td>
<td>28</td>
<td>3.52</td>
<td>3.36</td>
<td></td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>23</td>
<td>27</td>
<td>3.68</td>
<td>3.24</td>
<td></td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>24</td>
<td>26</td>
<td>3.84</td>
<td>3.12</td>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td>Equal distribution of high risk (like RCT)</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>4.00</td>
<td>3.00</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td>No. of excess high-risk patients in fibrinolysis</td>
<td>2</td>
<td>27</td>
<td>23</td>
<td>8.10</td>
<td>5.80</td>
<td></td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>28</td>
<td>22</td>
<td>8.40</td>
<td>5.54</td>
<td></td>
<td>0.72</td>
</tr>
</tbody>
</table>

PPCI indicates primary angioplasty; RCT, randomized, controlled trial.

In this cohort of 200 patients, an excess of only 6 high-risk patients in PPCI (3% of the total cohort) is sufficient to effectively abolish the benefit of PPCI. We modeled 1-year mortalities as follows: for low risk, 3.6% with PPCI and 4% with fibrinolysis; for higher risk, 25% with PPCI and 30% for fibrinolysis based on recent mortality data. These simple values were used so that PPCI would have 0.84 the mortality of fibrinolysis overall, with a relative risk reduction of 10% in the low-risk subgroup and 16% in the high-risk group (Killip >1), high-risk patients would have ≈7 times the mortality of low risk (for the same treatment arm). This is a conservative estimate given the timeline period over which the registries were performed (1991–2004) and overall mortality of the entire cohort would be ≈10%. This is consistent with published data.1,15–19,24

Results

Hypothetical Model

The results of the hypothetic model of registries allowed the prediction of the effect of unequal high-risk patient distribution between PPCI and fibrinolysis in the registries. The model suggested that mortality differences between PPCI and fibrinolysis would be very sensitive to high-risk patient distribution, with only 3 patients (equating to an excess prevalence of 6% in the PPCI arm) needing to be switched from fibrinolysis to PPCI for the benefit of PPCI to no longer be evident (Table 1 and Figure 1).

Analysis of Registries

In 8 of the 11 registries, there were sufficient data enumerating the higher-risk patients according to our definition (50 494 patients, 92% of the patients in the meta-analyses examined by Huynh et al13, Table 3). The remaining 3 registries, Drija et al10 (662 patients), Myocardial Infarction Triage and Intervention Registry (MITI)5 (3145 patients), and Unites des Soins Intensives Coronariens I (USIC I)6 (721 patients), did not publish sufficient data to be included in our analysis.

Allocation of High-risk Patients

In 6 of the 8 registries, more high-risk patients were allocated to PPCI. In the remaining 2, more high-risk patients were allocated to fibrinolysis (Table 3). We found the difference in mortality between PPCI and fibrinolysis to be strongly dependent on the unequal allocation of high-risk patients, with arms receiving excess high-risk patients tending to show worse long-term mortality; the difference in allocation of high-risk patients explained a large proportion of the variance between registries. When these covariates were included in separate meta-regression models, mortality differences between PPCI and fibrinolysis remained nonsignificant, although mean odds ratios tended to favor PPCI (but with confidence intervals that included 1) (Table 4).

Discussion

This study shows how sensitive STEMI registries are to expert clinician preference for therapy allocation in higher-risk patients. It is the first to quantify the impact of this effect, statistically called allocation bias, across the published...
registries. Adjustment for this bias shows PPCI to have similar significant mortality benefits over fibrinolysis in registries and in RCTs.

**Direction of Allocation Bias and Registry Mortality**

Our model demonstrated how outcomes of registries can be disproportionately influenced by a group of patients whose baseline characteristics are easily identified by the clinician before therapy and the influence of this bias in either direction. Most important, it predicted that only a small absolute number of such patients being preferentially allocated to 1 arm rather than the other is sufficient to distort results.

Real-world patients are not allocated randomly to therapies but rather are allocated to the therapy the physician deems most appropriate. Some STEMI patients such as those presenting with cardiogenic shock are obviously higher risk at presentation. PPCI is associated with lower (but still very substantial) mortality in such patients. The registry data in this analysis show that in some emergency systems, higher-risk patients preferentially undergo PPCI; in others, the tendency may be in the opposite direction with higher-risk patients undergoing fibrinolysis.

Because a disproportionate number of deaths in STEMI occur in higher-risk patients and the absolute difference in mortality between therapies is often small, only minor preferential allocation of higher-risk patients to 1 therapy can easily have a surprisingly large effect on relative mortality between therapies.

**Quantifying How Sensitive a Registry Is to Allocation Bias**

Although physician-led preferential allocation of patients in registries is recognized as a potential source of bias, our analysis is the first to use a simple criterion of higher risk across several registries to quantify the magnitude of this allocation

### Table 2. Sensitivity Analysis: Calculating the Number Need to Abolish for Different Therapeutic and Population Characteristics

<table>
<thead>
<tr>
<th>Relative Risk of PPCI Vs Fibrinolysis</th>
<th>Relative Risk Increase</th>
<th>Proportion of High Risk in STEMI Population</th>
<th>No. of Patients Per 100 That Need to Be Diverted From Fibrinolysis to PPCI to Abolish the Better Outcomes in the PPCI Arm (Number Needed to Abolish)</th>
<th>Excess Prevalence of High Risk Required in PPCI Arm for Benefit to Be Lost (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>8</td>
<td>0.20</td>
<td>2</td>
<td>3.42</td>
</tr>
<tr>
<td>0.90</td>
<td>8</td>
<td>0.25</td>
<td>2</td>
<td>3.95</td>
</tr>
<tr>
<td>0.90</td>
<td>8</td>
<td>0.30</td>
<td>2</td>
<td>4.47</td>
</tr>
<tr>
<td>0.90</td>
<td>6</td>
<td>0.20</td>
<td>2</td>
<td>4.39</td>
</tr>
<tr>
<td>0.90</td>
<td>6</td>
<td>0.25</td>
<td>2</td>
<td>4.91</td>
</tr>
<tr>
<td>0.90</td>
<td>4</td>
<td>0.20</td>
<td>2</td>
<td>4.74</td>
</tr>
<tr>
<td>0.90</td>
<td>4</td>
<td>0.25</td>
<td>3</td>
<td>5.26</td>
</tr>
<tr>
<td>0.90</td>
<td>4</td>
<td>0.30</td>
<td>3</td>
<td>5.79</td>
</tr>
<tr>
<td>0.84</td>
<td>8</td>
<td>0.20</td>
<td>3</td>
<td>5.65</td>
</tr>
<tr>
<td>0.84</td>
<td>8</td>
<td>0.25</td>
<td>3</td>
<td>6.52</td>
</tr>
<tr>
<td>0.84</td>
<td>8</td>
<td>0.30</td>
<td>3</td>
<td>7.39</td>
</tr>
<tr>
<td>0.84</td>
<td>6</td>
<td>0.20</td>
<td>3</td>
<td>6.38</td>
</tr>
<tr>
<td>0.84</td>
<td>6</td>
<td>0.25</td>
<td>4</td>
<td>7.25</td>
</tr>
<tr>
<td>0.84</td>
<td>6</td>
<td>0.30</td>
<td>4</td>
<td>8.12</td>
</tr>
<tr>
<td>0.84</td>
<td>4</td>
<td>0.20</td>
<td>4</td>
<td>7.83</td>
</tr>
<tr>
<td>0.84</td>
<td>4</td>
<td>0.25</td>
<td>4</td>
<td>8.70</td>
</tr>
<tr>
<td>0.84</td>
<td>4</td>
<td>0.30</td>
<td>5</td>
<td>9.57</td>
</tr>
<tr>
<td>0.60</td>
<td>8</td>
<td>0.20</td>
<td>8</td>
<td>16.25</td>
</tr>
<tr>
<td>0.60</td>
<td>8</td>
<td>0.25</td>
<td>9</td>
<td>18.75</td>
</tr>
<tr>
<td>0.60</td>
<td>8</td>
<td>0.30</td>
<td>11</td>
<td>21.25</td>
</tr>
<tr>
<td>0.60</td>
<td>6</td>
<td>0.20</td>
<td>9</td>
<td>18.33</td>
</tr>
<tr>
<td>0.60</td>
<td>6</td>
<td>0.25</td>
<td>10</td>
<td>20.83</td>
</tr>
<tr>
<td>0.60</td>
<td>6</td>
<td>0.30</td>
<td>12</td>
<td>23.33</td>
</tr>
<tr>
<td>0.60</td>
<td>4</td>
<td>0.20</td>
<td>11</td>
<td>22.50</td>
</tr>
<tr>
<td>0.60</td>
<td>4</td>
<td>0.25</td>
<td>13</td>
<td>25.00</td>
</tr>
<tr>
<td>0.60</td>
<td>4</td>
<td>0.30</td>
<td>14</td>
<td>27.50</td>
</tr>
</tbody>
</table>

PPCI indicates primary angioplasty; STEMI, ST-segment elevation myocardial infarction.

This table shows the excess proportion of higher-risk patients required to be allocated to PPCI to eliminate benefit at different relative risks between PPCI and fibrinolysis, different relative risk of high- vs low-risk patients, and differing proportion of high-risk patients in the study population. It can be seen that the number needed to abolish remains small across a spectrum of therapeutic and disease characteristics.
bias and to model the impact of a variety of potential sizes of bias on expected mortality.

The effect of allocation bias holds true for all clinical therapies and is particularly marked when baseline patient characteristics available at the time the clinical decision is made powerfully affect mortality (eg, evaluating the use of thrombolysis in stroke or activated protein C in sepsis). We have therefore derived an NNA formula to calculate the number of patients required to be allocated in a biased fashion to appear to destroy any benefit of a beneficial therapy (or, equally, to create an artifactual benefit in an ineffectual therapy). Evaluation of this NNA requires only the

Table 3. Registries According to Higher-Risk Patient Distribution and Long-term Mortality

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Patients</th>
<th>Marker of High Risk</th>
<th>Excess High-Risk Patients in PPCI Arm (A) (%)</th>
<th>Proportion of High-Risk Thrombolysis Patients (B) (%)</th>
<th>Proportion of High-Risk PPCI Patients (%)</th>
<th>Proportion of High-Risk Patients Allocated in Excess to PPCI (A/B)</th>
<th>Log Odds Ratio (95% Confidence Interval) (&lt;0 Means PPCI Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USIC II7</td>
<td>979</td>
<td>Killip &gt;1</td>
<td>5.3</td>
<td>13.1</td>
<td>18.4</td>
<td>0.40</td>
<td>0.12 (−0.29 to 0.53)</td>
</tr>
<tr>
<td>MISTRAL21</td>
<td>1585</td>
<td>Killip &gt;1</td>
<td>6.2</td>
<td>18.9</td>
<td>25.1</td>
<td>0.33</td>
<td>0.01 (−0.29 to 0.34)</td>
</tr>
<tr>
<td>HSU et al20</td>
<td>202</td>
<td>Killip &gt;1</td>
<td>8.4</td>
<td>31.4</td>
<td>39.8</td>
<td>0.27</td>
<td>−0.73 (−1.55 to 0.10)</td>
</tr>
<tr>
<td>De Labroille et al8</td>
<td>794</td>
<td>Killip &gt;1</td>
<td>2.9</td>
<td>15.8</td>
<td>18.7</td>
<td>0.18</td>
<td>−0.46 (−0.95 to 0.04)</td>
</tr>
<tr>
<td>ALABAMA</td>
<td>348</td>
<td>High-risk features</td>
<td>4.0</td>
<td>36.0</td>
<td>42.0</td>
<td>0.11</td>
<td>0.29 (−0.42 to 1.01)</td>
</tr>
<tr>
<td>CCP8</td>
<td>15 940</td>
<td>Killip &gt;1</td>
<td>1.4</td>
<td>36.7</td>
<td>38.1</td>
<td>0.04</td>
<td>−0.01 (−0.14 to 0.12)</td>
</tr>
<tr>
<td>ACOS12</td>
<td>4441</td>
<td>Cardiogenic shock</td>
<td>−1.8</td>
<td>6.7</td>
<td>4.9</td>
<td>−0.27</td>
<td>−0.31 (−0.61 to −0.01)</td>
</tr>
<tr>
<td>RIKS-HIA</td>
<td>26 205</td>
<td>Killip &gt;1</td>
<td>−10.3</td>
<td>30.1</td>
<td>19.8</td>
<td>−0.34</td>
<td>−0.76 (−0.86 to −0.66)</td>
</tr>
</tbody>
</table>

PPCI indicates primary angioplasty, USIC II, Unites des Soins Intensives Coronariens II; MISTRAL, Myocardial Infarction With Severe prognosis: observation of Treatment with Angioplasty or Lysis; ALABAMA, Alabama Registry of Myocardial Ischemia Investigators; CCP, Cooperative Cardiovascular Project; ACOS, Acute Coronary Syndromes; and RIKS-HIA, Register of Information and Knowledge about Swedish Heart Intensive Care Admissions.

The real-world registries included in the analysis according to high-risk patient distribution and long-term mortality.
relative risk reduction of the better therapy, the relative risk increase from being higher risk, and the proportion of such higher-risk patients in the population—information generally available from RCTs.

The 3 characteristics of a condition that make it sensitive to allocation bias are evident from the formula. If the higher-risk group is small (low $f_{high}$), very much higher in risk (low $1/RRI_{high}$), or the relative risk reduction from the better therapy is small (small $RRR$), then registries are especially vulnerable (low $NNA$). Calculating $NNA$ may therefore assist interpretation of registry data.

**Determining the Relative Merits of Different Reperfusion Therapies: The Role of Registries**

The detrimental effect of confounding is recognized, and attempts to quantify its effect on registry data in epidemiological studies have been proposed.27,28 Our findings are consistent with these foundational studies—namely that the effect of confounders is large and commonly underestimated. With the current prioritization of comparative efficacy research,29 it is important that clinicians appreciate the impact that confounders may have when they are interpreting comparative effectiveness research data. The effect of allocation bias has

---

**Figure 2.** Meta-regression of the relationship between differential allocation of higher-risk patients and odds ratio of mortality between fibrinolysis and primary angioplasty (PPCI) in selected registries. This figure illustrates the results of the meta-regression and confirms the predictions of our model. The therapy with the greater proportion of high-risk patients tends to have a higher-than-expected mortality. In fact, the high-risk covariate explains a significant proportion of the variance between studies ($R^2_{meta} = 83.1\%$). PPCI provided a significant 22% reduction in mortality compared with fibrinolysis when there was no bias in high-risk patient distribution (interpolated odds ratio, 0.78; confidence interval [CI], 0.64–0.97; $P=0.029$). The odds ratio for each registry is shown with 95% tolerance intervals (in RIKS-HIA and CCP the tolerance interval falls within the marker). The size of each marker is proportional to the sample size of each study, and the dashed lines represent hyperbolic uncertainty bounds. RIKS-HIA indicates Register of Information and Knowledge about Swedish Heart Intensive Care Admissions, CCP, Cooperative Cardiovascular Project.

**Table 4.** Effect of Inclusion of Various Covariates on the Estimated Odds Ratio for Fibrinolysis Versus PPCI in the Pooled Selected Observational Data

<table>
<thead>
<tr>
<th>Covariate Included in Model</th>
<th>Adjusted Mortality Odds Ratio (&lt;1 Means PPCI Better)</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td>0.79 (0.56–1.12)</td>
<td>0.19</td>
</tr>
<tr>
<td>High risk</td>
<td>0.78 (0.64–0.95)</td>
<td>0.02</td>
</tr>
<tr>
<td>Previous stroke</td>
<td>0.74 (0.47–1.17)</td>
<td>0.15</td>
</tr>
<tr>
<td>Female</td>
<td>0.80 (0.30–2.15)</td>
<td>0.60</td>
</tr>
<tr>
<td>Hypertension</td>
<td>0.72 (0.41–1.28)</td>
<td>0.21</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>0.79 (0.55–1.13)</td>
<td>0.16</td>
</tr>
</tbody>
</table>

$PPCI$ indicates primary angioplasty.

The table shows odds ratios after adjustment for each individual covariate in separate random-effects meta-regression models.
never previously been quantified in this field and therefore not accounted for in the published meta-analysis that found no significant benefit of PPCI.\textsuperscript{13} We have shown that the effect of allocation bias on mortality within registries is far larger than might be suspected. Once this allocation bias is accounted for, registries predict that PPCI has 22\% lower long-term mortality than fibrinolysis. This is consistent with the odds ratio favoring PPCI suggested by a recent meta-analysis of RCTs.\textsuperscript{13}

RCTs are difficult to conduct, are expensive, and may not be properly representative of patient populations. Registries, covering wide patient populations, are argued to be more generalizable because they can interrogate a therapy in subgroups excluded from RCTs\textsuperscript{14} and may be favored for reasons of ethics or feasibility. Registries and meta-analyses of them may therefore become more common. Although registries often incorporate adjustment for allocation bias in various ways, eg, by covariate adjustment or propensity scoring, they can adjust only for covariates or combinations of covariates that were documented. There are many features are obvious to a clinician at the time of decision making that may signify high risk—such as sweating, tachypnea, altered consciousness, low socioeconomic status, poor education, or the inability to comply with subsequent medical therapy—but are not routinely quantified and so rarely become available as covariates. All of this can overwhelm the best-adjusted analysis. Moreover, the variety of possible adjustments creates uncertainty in interpretation, and meta-analyses\textsuperscript{11} may use only the unadjusted mortality data to maintain methodological consistency across data sets. Because randomization accounts for known and unknown confounders, RCTs should remain the gold standard when determining the relative merits of competing interventions. However, although RCTs avoid problems of allocation bias, their strict inclusion criteria make generalization of their results to real-world practice difficult. One solution to this problem would be to design more pragmatic RCTs (with broad inclusion criteria and objective end points) to better mirror real-world practice.

**Study Limitations**

To limit the possibility of a false-positive result, we have used a random-effects analysis, but it remains unclear how the number of studies, the relative weights of each study, the extent of heterogeneity between studies, and the potential for aggregation bias all influence the probability of any meta-regression producing false-positive or false-negative results.\textsuperscript{26,30} Therefore, any conclusions from our analysis should be interpreted with this uncertainty in mind.

In each of the registries, we used the risk classification provided by the authors, which differed slightly between registries. However, in each registry, the criteria were the same between arms. Therefore, our findings are unimpaired that the registries each showed a tendency to refer more higher-risk patients to either PPCI or fibrinolysis and that this allocation bias powerfully affected reported mortality. Other markers of higher risk (such as PAMI [Primary Angioplasty in Myocardial Infarction], TIMI [Thrombolysis In Myocardial Infarction], or GRACE [Global Registry of Acute Cardiac Events] score) could not be used because none were routinely reported in the majority of registries. Had they been available, their broader base and finer quantitative grain may have permitted even more extensive adjustment.

Some studies did not make any attempt to account for possible confounders, although others that did adjusted for differing covariates, preventing the pooling of the adjusted odds ratios in the meta-analysis. It is conceivable that if all the registries had all been similarly adjusted, use of the adjusted outcomes would have been more appropriate, and perhaps this would not mask the benefit of PPCI. However, such a meta-analysis may be viewed with caution by the community because not every reader would be confident that the choice of which adjustment method to publish was unbiased and that the choice of which published adjusted values the meta-analysts chose to use was also unbiased. Therefore, we can never be certain that a meta-analysis of similarly adjusted registries will produce reliable results. An alternative would be to undertake an individual patient-level meta-analysis in which adjustments can be made across all studies. However, it is often difficult to get these data from the study investigators, and often the factors of interest have not been measured, or there is an unknown source of bias that cannot be adjusted for. It was presumably because of this concern that Huynh et al\textsuperscript{13} chose to use unadjusted values in their meta-analysis. To ensure methodological consistency and to enable pooled analysis across the studies, we adopted the same methodology as the Huynh et al\textsuperscript{13} analysis and used unadjusted outcomes and then adjusted for Killip Class.

**Conclusions**

The conflict between RCT and registry meta-analyses of PPCI versus fibrinolysis can be explained by allocation bias of just a few higher-risk patients to the PPCI arm in registries. We present a quantitative formula highlighting 2 factors that make registry studies comparing interventions vulnerable to bias: the ability of clinicians on the spot to identify a high-risk subset and a tendency to preferentially allocate them to 1 therapy. The principal challenge to comparative effectiveness research from an observational data set may be insuperable: the investigators can adjust only for the information that is documented systematically, which is only a subset of the information used by the clinician on the spot. Observational comparative effectiveness research is therefore led astray, not by poor research practice but by good medical practice.

Meaningful comparative effectiveness research requires the populations having the 2 therapies to be matched for not only documented characteristics but also undocumented ones. When the doctors making decisions are skilled, this is achievable only in the structure of a RCT.

**Sources of Funding**

Dr Sen is funded by the Medical Research Council UK (G1000357), Dr Davies (FS/05/006) and Dr Francis (FS10/038) are funded by the British Heart Foundation.

**Disclosures**

None.
References


Why Does Primary Angioplasty Not Work in Registries? Quantifying the Susceptibility of Real-World Comparative Effectiveness Data to Allocation Bias
Sayan Sen, Justin E. Davies, Iqbal S. Malik, Rodney A. Foale, Ghada W. Mikhail, Nearchos Hadjiloizou, Alun Hughes, Jamil Mayet and Darrel P. Francis

Circ Cardiovasc Qual Outcomes. published online November 13, 2012;
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/early/2012/11/13/CIRCOUTCOMES.112.966853

Data Supplement (unedited) at:
http://circoutcomes.ahajournals.org/content/suppl/2012/11/13/CIRCOUTCOMES.112.966853.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/
SUPPLEMENTAL MATERIAL
Derivation of a general formula to calculate how many patients need to be diverted to a better arm to appear to abolish its advantage (‘Number Needed to Abolish’)

It is possible to calculate a general formula for the number of higher risk patients per 100 total patients that need to be diverted from an intrinsically worse (“control”) arm to an intrinsically better (“therapy”) arm, for the benefits of the better arm to disappear (Number Needed to Abolish; NNA) that is independent of the size of the study and the mortality. What matters is:

- the difference between arms, expressed as a relative risk reduction, $RRR$;
- how much higher the risk of higher risk patients is, than low risk patients, expressed as a Relative Risk Increase, $RRI_{high}$;
- the proportion of patients who are higher risk, $f_{high}$.

The mortality for low-risk patients is $M$, for higher-risk patients it is $M(1 + RRI_{high})$, therefore in the control arm the overall mortality is $M(1 + RRI_{high} \cdot f_{high})$. In the therapy arm, the mortalities are influenced by the intervention so total mortality is $(1 - RRR)M(1 + RRI_{high} \cdot f_{high})$. If the allocation of higher-risk patients to the control arm is biased, the NNA, required to equalise mortalities equals:

$$M\left(1 + RRI_{high} \left(f_{high} - \text{NNA}\right)\right) = M\left(1 - RRR\right)\left(1 + RRI_{high} \left(f_{high} + \text{NNA}\right)\right)$$

Dividing by $M \cdot RRI_{high}$:

$$\frac{1}{RRI_{high}} + f_{high} - \text{NNA} = \left(1 - RRR\right)\left(\frac{1}{RRI_{high}} + f_{high} + \text{NNA}\right)$$

$$RRR\left(\frac{1}{RRI_{high}} + f_{high}\right) = (2 - RRR)\text{NNA}$$

Therefore, the number of patients that need to be switched to abolish any visible benefit is:

$$\text{NNA} = \frac{RRR}{2 - RRR} \left(\frac{1}{RRI_{high}} + f_{high}\right)$$

$RRI_{high}$ and $f_{high}$ are readily available to authors of RCTs or registries, and to readers (if published). $RRR$ may be estimated from RCTs. Inserting values into the general formula
rapidly calculates how easy or difficult it would be for biased allocation to apparently abolish
the therapy benefit in unadjusted results of real-world registries studying the same
therapeutic question if the underlying benefit were as suggested in the RCT. The converse is
also true, namely that biased allocation could create a false apparent benefit when none
exists.

For example in our simulation model at the beginning of this paper $RRR=0.16$, $RRI_{high}=6$
and $f_{high}=0.25$. From this we see that the Number Needed to Abolish is $(0.16/2-0.16)*(0.25+(1/6)) = 3$ per hundred patients, which accords with the simulation in Table 1.

In Table 2 we perform a sensitivity analysis and tabulate what NNA has to be for a range of
situations. It shows that NNA is small: a small proportion of the total number of patients in
the registry.