NBER WORKING PAPER SERIES

TRIALS AVOID HIGH RISK PATIENTS AND UNDERESTIMATE DRUG HARMS

Jason Abaluck
Leila Agha
Sachin Shah

Working Paper 34534
http://www.nber.org/papers/w34534

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2025

We are thankful to Daniel Park and Brendan Balthis for their expert research assistance. We are grateful for Christopher Manz and Hari Raman for their thoughtful guidance on how to classify oncology trials and identify trial-eligible cohorts in SEER-Medicare data. We thank David Chan and seminar participants at University of Massachusetts-Amherst for their helpful feedback on this project. Sachin Shah received grant funding from NIA K76AG074919 to support this work. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Trials Avoid High Risk Patients and Underestimate Drug Harms
Jason Abaluck, Leila Agha, and Sachin Shah
NBER Working Paper No. 34534
December 2025
JEL No. C93, I1, O38

ABSTRACT

The FDA does not formally regulate representativeness, but if trials under-enroll vulnerable patients, the resulting evidence may understate harm from drugs. We study the relationship between trial participation and the risk of drug-induced adverse events for cancer medications using data from the Surveillance, Epidemiology, and End Results Program linked to Medicare claims. Initiating treatment with a cancer drug increases the risk of hospitalization due to serious adverse events (SAE) by 2 percentage points per month (a 250% increase). Heterogeneity in SAE treatment effects can be predicted by patient's comorbidities, frailty, and demographic characteristics. Patients at the 90th percentile of the risk distribution experience a 2.5 times greater increase in SAEs after treatment initiation compared to patients at the 10th percentile of the risk distribution yet are 4 times less likely to enroll in trials. The predicted SAE treatment effects for the drug's target population are 15% larger than the predicted SAE treatment effects for trial enrollees, corresponding to 1 additional induced SAE hospitalization for every 25 patients per year of treatment. We formalize conditions under which regulating representativeness of SAE risk will lead to more externally valid trials, and we discuss how our results could inform regulatory requirements.

Jason Abaluck
Yale University
Yale School of Management
and NBER
jason.abaluck@yale.edu

Leila Agha
Harvard University
Harvard Medical School
Department of Health Care Policy
and NBER
agha@hcp.med.harvard.edu

Sachin Shah
Massachusetts General Hospital
and Harvard University.
sshah@mgh.harvard.edu

1 Introduction

Randomized controlled trials evaluating new drugs are intended to provide safety and efficacy evidence to inform regulatory and clinical decisions. But, if trial enrollees are not representative of the population that ultimately takes up treatment, trials may fail to generate externally valid findings needed to inform drug approval decisions and treatment choices. FDA efforts and academic evaluations of trial representation have focused on demographic dimensions (such as sex, race, and age), but treatment effects are more closely tied to clinical comorbidities and frailty than these broad demographic categories.[1]

In this paper, we study whether patients at high risk of drug-induced serious adverse events (SAE) are underrepresented in cancer drug trials. Serious adverse events (SAEs) are a crucial determinant of heterogeneity in the net benefits of drugs. Cancer drugs often have toxic side effects, such as dangerous infections and organ failure. These side effects limit the efficacy of treatment by forcing early discontinuation of treatment, preventing patients from receiving the therapeutic dose, and causing premature deaths (Muss et al., 2007; Hurria et al., 2011; Shah et al., 2020). Yet, current trials may underestimate these risks. Approximately one-third of all novel therapies approved between 2001 and 2010 had a post-market FDA safety event (i.e., boxed warning, FDA safety communication, or withdrawal) (Downing et al., 2017), raising questions about whether current approaches to trial regulation are adequate to surface relevant safety information at the time of drug approval.

To evaluate the external validity of drug safety findings in trials, we need to compare SAE treatment effects in trials to SAE treatment effects in the target population. Trials measure and report adverse events in ways that are often poorly documented, inconsistently defined across trials, and difficult to replicate in large administrative datasets (Sivendran et al., 2014), precluding direct comparison of SAE rates in trials vs. observational data. Trials also fail to collect systematic data on patient comorbidities and other important SAE risk factors. For these reasons, we develop our own estimates of how SAE rates vary depending on patient risk factors, using data where we

[1] For recent regulatory measures and consensus reports on trial representativeness, see: National Academies of Sciences and Medicine (2022); Food and Drug Administration Oncology Center of Excellence, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research (2022); **?**.

observe both the target population and trial participants.

Our analysis proceeds in three steps using data from the Surveillance, Epidemiology, and End Results (SEER) Program linked to Medicare claims, as well as trial information from ClinicalTrials.gov. First, we predict SAEs in the first three months following initiation of cancer drug treatment using patient comorbidities, frailty, and demographic characteristics. Second, we demonstrate that this model of SAE risk systematically predicts heterogeneity in SAE treatment effects, using a quasi-experimental approach leveraging staggered timing of treatment initiation after cancer diagnosis. Third, we find that trial-enrolled patients have lower predicted SAE treatment effects, compared to the target population.

To construct predictions of SAE risk, we identify a cohort of patients who initiate treatment with cancer drugs within 6 months of diagnosis and measure hospitalizations with a primary diagnosis code indicating a potential SAE. We predict each cancer patient's risk of hospitalization for a potential SAE after treatment initiation given their comorbidities, frailty, and age. We use a jackknife approach to develop SAE risk measures for each patient from a model that was estimated excluding data from the index patient.

Next, we exploit variation in the exact timing of treatment initiation relative to cancer diagnosis to estimate SAE treatment effects. Using an event study framework that compares treated individuals to not-yet-treated comparisons, we identify the effect of initiating treatment on SAE, controlling for time since cancer diagnosis, as well as the patient's cancer type and stage at diagnosis. Treatment initiation with a cancer drug leads to a large and sudden spike in the SAE rate, increasing a patient's monthly risk of hospitalization due to an SAE by 2 percentage points. We then relate treatment effects to (jackknifed) risk, finding that SAE treatment effects are much larger for higher-risk patients: patients in the top third of the risk distribution experience twice as large an increase in SAE rates as patients in the bottom third. Our measure of SAE risk thus predicts substantial heterogeneity in the effect of cancer drug treatment on SAE hospitalizations.

Following this, we identify trial participants in SEER-Medicare data and compare them to non-participants. This approach allows us to investigate trial selection patterns within a fixed subpopulation. We ascertain trial participation using a newly validated method that leverages

National Clinical Trial (NCT) identifiers reported in Medicare claims (Green et al., 2022b). A single trial generally tests a single drug for a target population and is intended to be informative about the treatment benefits and harms in that target population (not all off-label indications for which a drug might be prescribed). Thus, we aim to benchmark trial representation using the target population of patients stated by the trialists themselves. To identify each trial's target population, we match each trial participant to the public NCT record in the ClinicalTrials.gov database, and identify the cancer type, stage, treatment line, and relevant tumor markers targeted by the trial. Under the guidance of a practicing oncologist, we identify cohorts of patients in registry-linked claims data that matched the relevant trial's target population. The inclusion of cancer registry data allows us to identify stage at diagnosis and certain tumor markers, while claims data allow us to infer treatment line (i.e., first vs. second drug treatment regimen) and disease progression (based on subsequent treatment choices following diagnoses). Together, these data allow us to compare trial participants to non-participants who would be eligible for the treatment.

Finally, we compare SAE risk among trial enrollees to SAE risk among the broader set of patients eligible for each treatment being evaluated. Patients at the 10th percentile of the SAE risk distribution are 4 times more likely to be enrolled in a clinical trial compared to patients at the 90th percentile. These patterns suggest that clinical trials systematically underestimate the adverse event risks associated with tested treatments by under-enrolling patients whose comorbid conditions and frailty increase their risk of adverse events. On average, the target population has 15% higher predicted SAE treatment effects compared to trial enrollees. This estimate is likely to understate the full gap between trial enrollees and non-enrollees because we compare SAE risk only among Medicare beneficiaries aged 65+.[2] We also show that the FDA's current goal of increasing enrollment of older patients (focused on ages 75+) to improve age representation of trial enrollees would not be sufficient to close the gap (Food and Drug Administration Oncology Center of Excellence, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research, 2022). To adequately power trials to estimate SAE treatment effects in a high risk

[2]A majority of incident cancers occur in patients over 65 years old, but these patients are underrepresented in cancer trials, suggesting that clinicians and regulators have to extrapolate findings on (presumably lower-risk) under-65 trial enrollees to (presumably higher-risk) over-65 enrollees. Our results show that subgroup analyses of those aged 65+ are inadequate to addressing concerns about the generalizability of trials to older patients.

subgroup (defined as patients in the top 30% of the risk distribution among treatment-eligible patients), trial sample sizes would need to be twice as large under current enrollment patterns as with representative enrollment.

The external validity of trial-estimated SAE treatment effects is crucial for informing both FDA policy action and individual clinical decisions. Even when potential health benefits appear to exceed induced adverse events, doctors seek to accurately inform patients about the possibility of severe harm. The classical model of trial representation shows representative trials minimize the mean squared error of the estimated average treatment effects in the target population. We derive formal conditions under which regulation requiring representativeness in terms of SAE also improves representativeness in terms of net impacts, taking into account both the drug's therapeutic benefit and induced harms.

Our insights could inform a new regulatory framework aimed at improving trial representativeness. This would involve: 1) pre-trial reporting of the anticipated target population, 2) reporting of predicted risk of trial enrollees' adverse events relative to this benchmark population (using a risk model such as the one we develop here), and 3) post-marketing surveillance to assess the correspondence between the prespecified target population and use of the drug for the indication in question.

An extensive clinical literature has considered the differences between efficacy in trials and effectiveness in clinical practice, often identifying gaps without isolating the underlying reasons (Nordon et al., 2016; Visram et al., 2023; Agha et al., 2023). Recent research has begun to consider the role of trial enrollment patterns as a potential contributor to these gaps.[3] Our paper has three primary innovations compared to the existing clinical literature on trial enrollment and adverse events. First, we quantify a relationship between SAE risk and treatment effects rather than assuming that patients at higher risk of potential SAEs also experience more SAEs as a causal result of treatment. Second, we measure treatment effects in the same way for trial and non-trial enrollees (or predicted trial enrollees), enabling an "apples-to-apples" comparison.[4] Third, we

[3]See, for example: Fujimoto et al. (2023); Jain et al. (2021); Orcutt et al. (2025); Rivera-Caravaca et al. (2018); Hanlon et al. (2022); Steg et al. (2007); Kennedy-Martin et al. (2015).

[4]Our approach contrasts with studies that benchmark against outcomes reported in trials, which often have measurement differences with analogous "real-world" quantities. Identifying adverse events through chart review vs.

observe directly which patients are in trials rather than inferring this from stated exclusion rules. To our knowledge, existing studies address at most one of these three challenges, and typically with much smaller scope (e.g. studying one trial or a few hospitals). We show that trial enrollees have *predictably lower treatment effects* in ways that an informed regulator could assess before the trial is conducted, provided they had data on means of comorbidities in both the trial and the target population.

Recent research in economics has also considered problems related to the external validity of trials (Vivalt, 2020; Al-Ubaydli et al., 2017). Manski (2017) argues that external validity merits as much attention as internal validity when evaluating and applying medical research, given that treatment response is usually heterogeneous. In a different context, Allcott (2015) points out that selective participation in field experiments has important implications for external validity. Chassang and Kapon (2022) develop a set of best practices for improving external validity, including measuring rich covariates of trial participants. In the context we study (i.e., cancer drug trials), this best practice is rarely applied: patients' non-cancer comorbidities are not documented in trial results and only rarely collected in trial data repositories. Our efforts to surmount these obstacles shaped our specific empirical approach, but future analyses could be facilitated by richer data collection.[5]

Prior research on how financial incentives shape research investment suggests a potential explanation for the patterns we document here (Oostrom, 2024; Budish et al., 2015; Agha et al., 2022). Trial findings have large effects on prescribing decisions (McKibbin, 2023; Azoulay, 2002), implying that there are strong private incentives to produce favorable results. The patterns of nonrepresentative enrollment we have uncovered may arise if trialists want to minimize reported harms, as these adverse outcomes factor heavily into both FDA decisions (Lackey et al., 2021; Tracy, 2025)

administrative records often yields divergent results (Connolly et al., 2024). EHR and claims data do not consistently capture the patient symptoms which may signify adverse events (Nadkarni, 2010). Even when comparing results for the same clinical trial reported in different forums (ClinicalTrials.gov vs. academic publications), adverse event rates often fail to match (Tang et al., 2015; Hughes et al., 2014).

[5]Research in biostatistics has developed transportability methods for estimating externally valid treatment effects, by reweighting trial observations to match the distribution of observable characteristics in a target population, or with doubly robust methods that also estimate treatment effect heterogeneity within trial data (Dahabreh et al., 2020, 2021; Degtiar and Rose, 2023; Robertson et al., 2024). These methods are poorly suited to our retrospective analysis of external validity, both because most trials did not systematically collect data on important predictors of SAE risk and because trials may rarely or never enroll patients at greatest risk.

and treatment choices (Busch et al., 2014; Chintagunta et al., 2009).

Improving the representativeness of trials could have important effects on which drugs are developed and which patients choose to use them. Kao and Oostrom (2024) and Michelman and Msall (2024) find that increasing clinical trial enrollment of older patients and women, respectively, fosters development of drugs to serve those populations. Alsan et al. (2024) studies the downstream consequences of nonrepresentative trials, finding that the underrepresentation of Black patients in clinical trials reduces their adoption of new drugs.

Trial findings are also a crucial input to FDA decisions. Cohen et al. (2021) and Olson (2002, 2008) found that shortening the timeline for FDA review led to the approval of drugs with worse safety profiles, while Philipson and Sun (2008) and Philipson et al. (2008) argue that FDA policy emphasizes safety at the expense of speed in approval decisions. By regulating representativeness of SAE risk among trial participants, it may be possible to surface safety information more quickly and accurately (with greater statistical power in smaller samples), lessening the potential tradeoff between speed and decision quality.

The paper proceeds as follows. Section 2 presents a model situating our focus within a conventional theory of trial representativeness. Section 3 discusses the context and data. Section 4 describes methods and results estimating heterogeneous effects of cancer drug initiation on SAEs. Section 5 investigates the relationship between SAEs and trial participation. Section 6 discusses a potential regulatory framework for improving representativeness and models how trial sponsors may respond. Section 7 concludes.

2 Formal Framework

This section addresses why externally valid estimates of serious adverse event (SAE) harms matter to clinicians, patients and regulators. We will also develop a formal framework that we will use in Section 6 to suggest an approach to regulating trial representativeness that would improve the external validity of severe adverse event effects.

Let $\tau_1, \ldots, \tau_N$ denote net impacts of treatment in subgroups $j = 1, \ldots, N$ which comprise respectively fractions $p_1, \ldots, p_N$ of the total population who will be treated with a drug (should

it be approved). Since trials are not typically powered for subgroup analysis, doctors learn only the estimated average treatment effect from trials (Alosh et al., 2015). Let r_j denote the fraction of each group j in a given trial. A simple model of trial design would imply that trialists seek to minimize the mean-squared error between $\hat{\tau}_{ATE}$ estimated from a trial and $\tau_{ATE} = \sum_j p_j \tau_j$, the true average treatment effect in the population. Classic results imply that, if trials are sufficiently large or all groups have equal variance of treatment effects, the mean-squared error minimizing trial design would represent all groups proportionately (Neyman, 1934; Cochran, 1977), setting $r_j = p_j$.

In practice, trials are conducted via convenience sampling, and so $r_j \neq p_j$. Participants are recruited from a relatively small set of (non-randomly selected) trial sites, with physicians exercising discretion to make enrollment offers and patients endogenously choosing whether to participate. Because randomization occurs after this enrollment process is completed, trials have strong internal validity. But trials' external validity relies on how representative treatment effects are for trial participants relative to the target population, given that participants are not a simple random sample of the target population.

Suppose that the net impacts of a drug for patient subgroup j can be written as $\tau_j = \alpha_j + \delta\gamma_j$, where α_j denotes treatment effects for intended medical benefits and γ_j denotes treatment effects for harms (i.e. adverse events), which have a relative welfare weight $\delta > 0$. We normalize the signs of each component so that α_j is negative when the patient receives beneficial effects of treatment (e.g. through reduced probability of mortality), and γ_j is positive when patients experience harms of treatment (e.g. through increased risk of organ failure).

We define $B_\gamma = \sum_j (r_j - p_j)\gamma_j$ as the bias with respect to severe adverse events—estimating this object will be the focus of most of our analysis. We focus on SAE representativeness rather than net impact representativeness, in part because our quasi-experimental design is not well-suited to evaluate health benefits that materialize over several months or years. A proportional trial could obtain unbiased estimates of harms $E(\gamma_j)$ and benefits $E(\alpha_j)$ separately. Lack of proportionality with respect to γ_j (i.e. $B_\gamma \neq 0$) is thus diagnostic that a trial is not setting $r_j = p_j$. However, one could conduct a similar test using any observable characteristic. Our question in this section is, "Why do we care about representativeness with respect to γ_j specifically, rather than any other

arbitrary patient characteristic (e.g. people with last names starting with K)?"

We provide two justifications. A first answer is that doctors, patients, and regulators may care about average harm $E_p(\gamma_j)$ in its own right separately from average net benefit $E_p(\tau_j)$ (where the p subscript denotes population expectations). If a cancer treatment is likely to prolong life expectancy on average but raises the risk of a debilitating stroke, patients want to be informed about the latter risk. We may thus want to design trials that are representative in terms of both τ_j and γ_j.

A second answer is that, even if doctors or regulators care only about net impacts τ_j, trials which are less biased in terms of average harm $\overline{\gamma}$ will also tend to be less biased in terms of average net impacts $\overline{\tau} = E_p(\tau)$, under conditions we make precise. Setting $r_j = p_j$ (i.e., trial is a random sample of the population) would lead to unbiasedness of both harms and net impacts, but there are other ways of oversampling some groups and undersampling others to get unbiasedness in terms of harms $\overline{\gamma}$ but not net impacts $\overline{\tau}$. In Section 6, we derive conditions under which trialists seeking representativeness in terms of severe adverse events in the least costly possible way will also improve net impacts.

3 Context & Data

3.1 Clinical Context: Generalizability of Cancer Trials

The external validity of cancer trials is a particularly salient problem for the care of older adults with cancer. More than 50% of incident cancers occur in people over 65 years old (National Cancer Institute, 2025). High rates of comorbid disease among older adults complicate treatment choices; over two-thirds of Medicare beneficiaries have two or more chronic conditions and more than one-third have 4 or more conditions. Concerns about drug safety, often framed as tolerability, are a major barrier to treating older adults with cancer, who are less likely to initiate treatment, more likely to discontinue treatment early, and more likely to experience serious adverse events due to chemotherapy (Bradley et al., 2008; Dy et al., 2006; Luo et al., 2006; Pasetto et al., 2008; O'Grady et al., 2011; Vinod et al., 2010; Neugut et al., 2006; Grønberg et al., 2010; Hassett et al., 2011;

Zauderer et al., 2009).

Concerns about trial generalizability have been a long-standing concern of both physicians and regulators. In 2020, FDA guidance established that clinical trials should be designed to be "representative of the population that will use the drug if approved" (U.S. Food & Drug Administration, 2020). This guidance was further operationalized in June 2024 when the FDA released draft guidance requiring trial sponsors to submit goal patient enrollment disaggregated by age group, sex, and race and ethnicity (U.S. Food & Drug Administration, 2024). Stated goals must be justified, for example, based on the disease distribution by demographic group. However, demographic variables, including age, race, and sex, do not capture many of the most clinically-relevant considerations. Comorbidities and geriatric syndromes (e.g., frailty, disability) are key features that influence physician decision-making (Sarfati et al., 2016; George et al., 2021). The current regulatory approach has neglected to provide guidance for measuring or ensuring that trials represent the distribution of clinical risk factors among patients who would use the drug.

3.2 SEER-Medicare data

This project uses data from the SEER Program (2014-2019) linked to Medicare claims (2013-2020) to study trial enrollment and generalizability. The SEER Program pools data from 16 state and regional cancer registries, spanning many large states including California, New York and Texas (see Appendix Table A1). These data capture all incident cancer diagnoses for 7 cancer types (bladder, breast, colon, lung, pancreas, rectal, renal). Using these data, we developed two distinct cohorts for the two phases of our investigation: first, to identify patients at highest risk of experiencing a serious adverse event (SAE) upon initiation of cancer drug treatment and to estimate heterogeneous SAE treatment effects; second, to identify how trial enrollment varies with predicted SAE treatment effects. See Appendix A2 for further details on cohort construction.

Adverse Event Sample. The first cohort was developed to estimate heterogeneous effects of cancer drug initiation on SAE risk. The cohort includes patients with new diagnoses of cancer who initiate drug treatment within 3–6 months following diagnosis. Initiation of cancer drug treatment is identified using the first cancer drug billed to Part B or Part D claims, following the drug definitions

described in Appendix A2. We exclude patients who initiate drug treatment immediately upon diagnosis, because hospitalizations ramp up around the time of diagnosis, including the month or two before diagnosis; for these patients, it is difficult to disentangle causal effect of initiating a new drug from the effects of developing cancer. (We discuss our approach to identification further in Section 4.2.) Using SEER data, we record patient's cancer type and stage at diagnosis. We require that patients be continuously enrolled in Medicare Parts A/B/D for 1 year before and after their diagnosis date (or until death, if they die within 1 year).

Serious adverse events are measured using diagnosis codes identified by prior researchers and reviewed by clinicians on our team.[6] We require a relevant code to appear as the primary diagnosis on an inpatient hospitalization claim. Our SAE definition spans several key types of serious adverse events associated with cancer treatment: infection, kidney, cardiac, hematologic, gastrointenstinal, liver and lung events. We measure SAE incidence at the patient-month level, defined as whether the patient had any SAE hospitalization during the month. In our baseline analysis, we retain patients in the sample even after death, and record no adverse events in months after death; in Appendix Figure A2, we show very similar findings with an alternative approach that drops patient-month observations after the patient's death. We also construct patient-level risk factors for adverse events as of the date of cancer diagnosis, including chronic conditions (as defined by the Chronic Condition Warehouse), a claims-based frailty index (as developed by Kim et al. (2018) to predict death, disability, recurrent falls, and health care utilization), age and sex.

Trial Participation Sample. The second cohort was developed to compare the SAE risk of trial participants to the full population of patients with the targeted disease, using the same SEER-Medicare data from 2014-2019. Since 2014, Medicare has required reporting of National Clinical Trial (NCT) database identifiers in claims for trial-related services. We use this reporting to identify trial participants in Medicare claims data, and link to the specific NCT record to identify the trial's eligibility group. Prior research has validated the use of these NCT records to identify participants in cancer trials (Green et al., 2022a). Following methods developed in Shah et al.

[6]See Bishnoi et al. (2021); Bittoni et al. (2018); Du et al. (2005); Freedman et al. (2014); Gunturu et al. (2022); Hansen et al. (2014); Herbach et al. (2022); Hershman et al. (2013); Rashid et al. (2015); Reeder-Hayes et al. (2017); Sanoff et al. (2012); Wieder and Adam (2023).

(2025), we used each trial's clinicaltrials.gov record to limit to randomized trials that tested a drug or biological agent. Then, we hired an oncologist who manually reviewed each trial with 5 or more SEER-Medicare participants; the physician identified the trial's enrollment criteria, including the targeted cancer subtype, stage, and treatment line. This review excluded trials that were not intended to treat the cancer (e.g., an RCT to prevent anemia) or were not specific to a single cancer (e.g., an RCT testing a drug that treats multiple types of solid tumors at once).

The Trial Participation Sample is used to make across-patient comparisons of the characteristics of trial participants and non-participants, and thus requires us to build a comparable sample of non-trial-participants who would potentially be eligible for the treatment tested in the trial. To support this analysis, we developed narrow subcohorts of patients that closely matched the target population specified in each trial's inclusion criteria. These subcohorts are defined by the primary cancer, relevant tumor markers, stage, and treatment line, e.g. second-line treatment of advanced-stage non-small cell lung cancer. These cohort definitions were developed with the guidance of an expert oncologist, based on clinically-relevant treatment categories. The patient's primary cancer, genetic tumor markers and stage at diagnosis are observed in SEER data. The treatment line is defined by how many prior treatment regimens the patient has had to date (so a second-line treatment is a treatment given only after a first therapeutic regimen failed), and we coded these based on observed drug treatment regimens in Medicare claims. In some cases, we also infer cancer stage has advanced beyond the stage at diagnosis based on the patient's treatment regimen.

For advanced stage cancers, we construct a target population of patients receiving active drug treatment. Because the drug treatments we leverage for cohort identification are specifically targeted to advanced cancers, this cohort captures both patients with advanced stage at diagnosis and those with disease progression to advanced stage. By contrast, there are multiple treatment strategies for early stage cancers, many of which do not use drug treatment. As a result, we construct a target population of patients who were diagnosed at an early stage regardless of treatment status. Given the differences in cohort construction, we report results for early vs. advanced stage cohorts separately as a complement to our main, pooled analysis.

This detailed mapping of SEER-Medicare data to clinically-relevant treatment cohorts was not

feasible for every disease state; we constructed 17 subcohorts spanning 4 common tumor sites (breast, lung, pancreatic and renal), and limit this analysis to these subcohorts. For more details on subcohort construction and definitions of the index date, see Appendix Table A2. For each patient in this cohort, we use 1-year claims history to measure the patient's comorbidities and frailty index, as well as sex and age.

3.3 Summary statistics

Summary statistics for the Adverse Event Sample are reported in Table 1, column 1. The average patient age is 74; 66% of patients are female; and 84% are White. Patients experience significant SAE risk after treatment initiation, with 4.6% of patients experiencing at least one hospitalization with primary diagnosis corresponding to an SAE within 90 days of treatment initiation. Appendix Table A3 further describes the frequency of each SAE category. The most common type of SAE is infection, comprising just over half of the total events. Chemotherapy can affect the immune system, increasing vulnerability to infection (e.g., pneumonia). The second most common category is kidney injury; nephrotoxicity is another common side effect of cancer drugs. Cardiac, hematologic and gastrointestinal complications each lead to hospitalizations for 0.3–0.4% of treated patients within 3 months, while hospitalization for liver and lung adverse events are rare.

Table 1 columns 2 and 3 report summary statistics for the trial participation cohort. Our sample includes 852 trial participants from 112 distinct trials. Rates of common comorbidities are lower among trial participants compared to non-participants. In Table 2, we show further evidence that trial participants are healthier than non participants, after controlling for disease state fixed effects (i.e. cancer type $\times$ stage at diagnosis $\times$ treatment line). The estimates indicate that trial participants are less likely to have each of the 20 measured comorbidities and have lower values of the Kim Frailty index (Kim et al., 2018). After Bonferroni correction for multiple comparisons, 15 out of 21 comorbidities have significantly lower incidence among trial participants at the 5% level. Trial participants are 3 years younger on average than non-participants (among our Medicare sample aged 65+). This evidence suggests that there is a selection pattern of healthier patients for trial participation. In the analysis that follows, we will assess what these differences imply for the

external validity of trial findings.

Returning to the the Adverse Event Sample, Figure 1 illustrates our identifying variation: this plot shows the average monthly SAE rate before and after cancer diagnosis. The plot separates patients into four groups based on the timing of treatment initiation: 3, 4, 5, or 6 months after diagnosis. In all four groups, SAE hospitalizations spike up initially around the month of cancer diagnosis. This pattern could occur because cancer itself causes worsening health and increases the hospitalization rate for a variety of health risks. The hospitalization rate falls in months 1 and 2 after diagnosis, but spikes up sharply in the month of drug treatment initiation. Each group reaches its peak SAE hospitalization rate in the month of treatment initiation. This graph illustrates why we exclude patients who initiate treatment within 2 months of diagnosis from our estimation sample: for these patients, it is difficult to disentangle the health event of cancer diagnosis from the effects of treatment initiation. It also shows that patients who initiate treatment in months 3–6 after diagnosis have similar hospitalization rates over the pre-period, bolstering confidence that these groups are comparable and evolving on parallel trends.

4 Serious Adverse Event Treatment Effects

We start by describing how we construct SAE risk, identifying the attributes which predict adverse events among treated beneficiaries. Second, we relate adverse event risk to adverse event treatment effects within a difference-in-difference framework.

4.1 Estimating SAE Risk

Methods. We use a LASSO regression to predict the risk of SAEs following treatment initiation. The outcome variable for this analysis is the average of three monthly indicators for having any SAE hospitalization, calculated in the month of treatment initiation and two months following. We denote this outcome SAE_{ic} for the SAE rate of patient i with cancer type and stage c. Because this outcome requires three months of post-treatment follow-up, we exclude patients who initiated treatment after February 2020 (who would not have a complete three-month window before the end of our data in May 2020).

The estimating equation is as follows:

$$SAE_{ic} = \beta X_i + \phi_c + \varepsilon_i \text{ subject to } \Sigma|\beta_j| \leq t \tag{1}$$

The LASSO predictors X_i include indicator variables for 21 comorbidities, total comorbidity count, Kim frailty index (Kim et al., 2018), age (linear and quadratic terms), and sex. All comorbidities are measured at the time of cancer diagnosis (strictly prior to beginning any cancer treatment). The LASSO also includes fixed effects ϕ_c for 20 interactions of cancer type and stage at diagnosis; these control variables are not subjected to LASSO regularization.

We estimate this regression in 100-fold cross-validation, leaving out 1% of the sample at a time, to obtain a patient-specific prediction of SAE risk that is calculated from a sample that excludes the index patient. To construct a risk measure that predicts variation in SAE risk within a trial targeting a given cancer type, we exclude cancer type by stage variation from the predicted risk index. Specifically, we compute: $Risk_i = \hat{\beta}_{-i} X_i + \overline{\phi}$ where $\overline{\phi}$ is a constant parameter that centers the predicted SAE rates around the overall mean SAE rate in the sample.

Results. Table 3 reports the results of our SAE risk prediction. To summarize the 100 separate hold-1%-out LASSO models, we report results from a single model estimated on the pooled data. The LASSO prediction retains 17 out of 21 comorbidities with nonzero coefficients, and also retains the total number of comorbidities, the Kim frailty index, age^2, and an indicator variable for whether the patient is female. The two most important predictors are the Kim frailty index and chronic kidney disease. A one standard deviation increase in either variable predicts a 0.5 percentage point increase in SAE risk, from a mean of 3.3% risk per month. Using the hold-one-out LASSO model to predict SAE risk for each patient ($Risk_i$), the resulting distribution indicates wide dispersion in risk across treated patients, with the 10th percentile patient having a 1.3% chance of an SAE each month, while the 90th percentile patient's risk is over 4 times larger at 6.2% per month.

4.2 Estimating SAE Treatment Effects

Methods. The variation in SAE risk estimated above could reflect both differences in treatment effects that depend on patient covariates and differences in baseline risk of hospitalizations for

SAE conditions. In the second stage of analysis, we estimate the effect of treatment initiation on SAE hospitalizations and test whether these effects are larger among patients with higher SAE risk. We use two alternative methods for this analysis, drawing on recent two-stage approaches to event study analysis that address variation in treatment timing and possible treatment effect heterogeneity (Borusyak et al., 2024; Gardner, 2022). The main goal of this analysis is to estimate the change in the SAE hospitalization rate before vs. after treatment initiation.

Our analytic approach isolates comparisons of patients who recently initiated treatment (0, 1, or 2 months ago) to patients who are about to initiate treatment (1, 2, or 3 months in the future). Given that SAE rates increase around the time of cancer diagnosis (even in the absence of drug treatment), we control for months relative to cancer diagnosis in all of our regression analyses.

The first approach models SAE outcomes SAE_{igpt} for patient i who is in treatment group g, p months since diagnosis in calendar month t as follows:

$$E(SAE_{igpt}) = \theta_0 Treat_{it} + \theta_1(Risk_i - \overline{Risk})Treat_{it} + \xi_g + \delta_p + \gamma_t + W'_{it}\beta \tag{2}$$

In equation 2, $Treat_{it}$ is a binary variable indicating that patient i received cancer drug treatment in month t. θ_0 is the average effect of treatment, and θ_1 allows the treatment effects to vary with each patient's SAE risk (calculated as a function of their comorbidities, frailty, and demographics at the time of cancer diagnosis, as described above). ξ_g captures treatment group fixed effects: treatment groups are defined by the number of months since diagnosis when the patient initiates treatment. (In our SAE sample there are four treatment groups, corresponding to treatment initiation dates 3, 4, 5, and 6 months after diagnosis.) δ_p captures health evolution before and after diagnosis, by including fixed effects for each month before and after diagnosis (from -12 to +6). γ_t captures calendar time fixed effects to control for time trends and seasonality. Finally, W'_{it} includes patient characteristics, including cancer type by stage fixed effects, patient age, and predicted SAE risk.

We estimate this equation in two stages, using an imputation estimator adapted from Borusyak et al. (2024).

1. We restrict the sample to untreated observations only. Because all observations in our sample

are treated within 6 months after cancer diagnosis, this restriction limits to patients who are not-yet-treated. Within the not-yet-treated sample, we estimate:

$$Y_{igpt} = \xi_g + \delta_p + \gamma_t + W'_{it}\beta + \varepsilon_{igpt} \tag{3}$$

2. For each treated observation in months 0, 1 or 2 after treatment initiation, we use our estimated coefficients from step 1 to impute SAE outcomes in the post-period in the absence of treatment. We then subtract imputed outcomes (assuming no treatment) from observed outcomes (for treated individuals). Specifically, we calculate:

$$\hat{\theta}_{igpt} = Y_{igpt} - \hat{\xi}_g - \hat{\delta}_p - \hat{\gamma}_t - W'_{it}\hat{\beta} \tag{4}$$

Our aggregate treatment effect estimate is a simple average of these $\hat{\theta}_{igpt}$. To test for heterogeneity in treatment effects for patients with different predicted SAE risk, we estimate the following regression:

$$\hat{\theta}_{igpt} = \theta_0 + \theta_1(Risk_i - \overline{Risk}) + \varepsilon_{igpt}. \tag{5}$$

Recall that $Risk_i$ is predicted as a function of the patient's comorbidities, frailty, and demographics, using the coefficients from a LASSO model estimated in a sample that excluded the index observation.

For all results estimating SAE effects, confidence intervals are calculated using a bootstrap procedure with 500 iterations, clustered at the patient level.

We also estimated an extended version of this regression model that allows us to plot event study coefficients before and after treatment initiation. To do so, we modify the first stage of the estimation procedure as follows. Using only untreated observations (from periods prior to treatment), we estimate the following equation:

$$Y_{gpit} = R'_{it}\theta^R + \xi_g + \delta_p + \gamma_t + W'_{it}\beta + \varepsilon_{igpt} \tag{6}$$

where R'_{it} includes a vector of fixed effects for months -6 through -1 prior to treatment initiation

and θ^R captures the effect of being R months from treatment initiation. We use months more than 6 months prior to treatment initiation as the reference period. To estimate post-period coefficients, we average the estimates $\hat{\theta}_{igpt}$ from Step 2 within each post-treatment month (0, 1 or 2).

Results. Figure 2 displays the results of the event study framework. The pre-period shows no statistically significant effects, bolstering the parallel trends assumption for SAE risk in early- and later-treated groups. SAE risk increases in the month of treatment initiation and the 2 months immediately afterwards, by 2 to 3 percentage points per month. Note that we can only identify 3 months of post-treatment initiation effects in our current sample, which limits to patients who initiated treatment within 3-6 months from diagnosis.[7] The descriptive evidence plotted in Figure 1 suggests that this 3-month post-period covers the months with peak SAE risk.

Table 4 summarizes our difference-in-differences findings, and formally tests whether our measure of SAE risk is predictive of heterogeneity in the impact of initiating cancer drug treatment. Column 1 reports estimates from our baseline specification that controls for calendar time, time since diagnosis, SAE risk, age, cancer type X stage, and fixed effects for cohorts defined by the timing of treatment initiation. On average, over the three calendar months following treatment initiation, the risk of SAE increases by 2.4 percentage points (95% CI: 2.1 to 2.7 p.p.) from a pre-treatment baseline of 0.96 percentage points, representing a 250% increase.

Table 4 Column 2 reports estimates where we test for heterogeneity in SAE treatment effects, depending on patients' predicted SAE risk (as estimated by the jackknifed LASSO model). We find that patients whose comorbidities, frailty, age and sex predict higher SAE risk also experience larger increases in SAE hospitalizations after treatment initiation: roughly half of the observed variation in SAE risk translates into differences in SAE treatment effects.

Figure 3 displays a binned scatterplot showing the relationship between SAE risk and patients' estimated individual treatment effect. The scatterplot shows a strong linear relationship, with higher SAE risk corresponding to larger increases in SAE rates after treatment initiation. Taken together, these results imply that cancer drugs substantially increase patients' likelihood of hospi-

[7] By month 6, all patients have initiated treatment, so we do not have a not-yet-treated group to identify time-since-diagnosis effects for month 6. This means the longest treatment horizon for which we can estimate a treatment effect is 2 months post-treatment initiation, which takes place 5 months from diagnosis for the month 3 treatment initiation group.

talization for a potential SAE, and these risks vary systematically and predictably across patients.

Robustness. We assess the robustness of these results to several alternative specifications, varying the regression model, estimation equation, and coding of patient deaths. Table 4 columns 3 and 4 assess the robustness of our SAE findings to an alternative specification that includes patient-level fixed effects. The baseline specification and this alternative yield remarkably similar estimates of both the average treatment effect and treatment effect heterogeneity as a function of patients' SAE risk.

Appendix Figure A1 and Appendix Table A4 report results from an alternative sample, expanded to include patients who initiate drug treatment between 3-11 months after diagnosis (rather than 3-6 months, as in our baseline sample). Results are extremely similar across both samples, although the broader sample show small but statistically significant pre-trends in the two months preceding treatment initiation. We interpret this as evidence that patients who initiate treatment much later after diagnosis may have a different evolution in their baseline SAE risk, and thus we prefer comparing patients whose treatment initiation dates are relatively close together (as in the baseline sample). An advantage of the broader sample is that it allows us to estimate additional post-period effects. We find that SAE rates trend down beginning 3 months after treatment initiation, although treatment discontinuation (often spurred by adverse events) may contribute to this downward trend.

Appendix Table A7 provides suggestive evidence on the external validity of our findings to patients with earlier treatment initiation dates, between months 0-2 following diagnosis. Using our model of SAE risk (reported in Table 3, we predicted SAE risk for each subcohort defined by the month of treatment initiation (relative to diagnosis). We do not find a strong relationship between the timing of treatment initiation and predicted SAE risk, with average monthly risk ranging from 3.4 to 4.1%. This suggests that our sample of patients who initiate treatment between 3–6 months after diagnosis may have similar SAE risk to earlier treated groups.

Finally, we test an alternative approach to handling patient deaths. Our baseline analysis includes data for 3 months post-treatment for all patients; if a patient dies after treatment initiation but before 3 months have elapsed, we record their SAEs in subsequent months as zeros. In Fig-

ure A2, we show results from an alternative specification where we censor observations from months following the patient's death. This alternative specification generates similar, though slightly larger treatment effect estimates in months 1 and 2 following treatment initiation.

5 Serious Adverse Events and Trial Participation

5.1 Econometric approach

The second phase of analysis investigates the relationship between SAE risk and trial participation. Recall that the Trial Participation Sample is limited to specific subgroups that we can match precisely to trial target populations, based on the criteria outlined in detail in Appendix Table A2. We use the SAE risk prediction model constructed in the Serious Adverse Event Sample to predict patient-level SAE risk in the Trial Participation Sample. To transport the SAE risk prediction model to the Trial Participation sample, we use patient-level risk factors and coefficients recovered from our LASSO regression described above: $Risk_i = \hat{\beta} X_i + \overline{\phi}$. Because we do not observe the average SAE rate in the Trial Participation Sample, we center these predictions around the mean SAE rate in the SAE sample.[8]

Next, we test whether our measure of SAE risk predicts differences in trial enrollment probability, conditional on the patient's disease state (cancer type, tumor markers, disease stage, and treatment line). We estimate linear regressions to understand the relationship between trial participation and SAE risk:

$$Trial_{is} = \alpha Risk_i + \phi_s + \varepsilon_i. \tag{7}$$

The main coefficient of interest is α, which multiplies the patient-level SAE risk measure. We control for disease state fixed effects ϕ_s where disease state references the subcohorts defined by cancer type, stage, genetic markers, and treatment line (as summarized in Appendix Table A2).

[8]Recall that the Trial Participation Sample includes some early stage cancer patients who are not currently treated with cancer drugs, and as a result do not experience drug-induced serious adverse events. We use this choice for these early stage cohorts because there are no widely accepted drug treatment options for these early stage cancers currently.

In some specifications we also control for patient-level demographic characteristics: age (linear and quadratic terms), sex, race, and ethnicity.

Finally, we compare the distribution of SAE risk across trial participants and non-participants. To ensure differences between the two cohorts are not driven by variation in the trial participation rate across disease states, we reweight each observation in the full population (which includes both participants and non-participants) to match the disease-state distribution among the restricted sample of trial participants only (Barsky et al., 2002; Chandra et al., 2024). Specifically, if s indexes disease state cohorts, then we reweight each observation by:

$$w_s = \frac{P(Trial_s)/P(trial)}{(1 - P(trial_s))/(1 - P(trial))}, \tag{8}$$

where $P(Trial_s)$ denotes the trial participation rate among patients with disease state s and $P(trial)$ denotes the trial participation rate in the pooled sample. After reweighting, we construct kernel density plots comparing the distribution of SAE risk in the full population to risk among trial participants, and we also compare the average SAE risk. Finally, to assess the implications of these differences in SAE risk for reported SAE treatment effects, we use the estimates from Equation 5 to construct predictions of SAE treatment effects for each patient (given their SAE risk $Risk_i$): $\hat{\theta}_i = \hat{\theta}_0 + \hat{\theta}_1(Risk_i - \overline{Risk})$. We compare the average SAE treatment effect among trial participants versus among all patients (in the reweighted sample).

5.2 Results: SAE Risk and Trial Participation

Figure 4 shows a binned scatterplot of the relationship between trial participation and SAE risk, controlling for disease state. There is a strong, downward-sloping relationship between trial participation and SAE risk, over the whole range of observed SAE risk. Table 5 column 1 reports the corresponding regression results, confirming that this relationship is statistically significant at the 1% level. These results imply that a patient at the 10th percentile of the SAE risk distribution would have a 1.36 percentage point increase in their monthly risk of an SAE upon initiating a cancer drug and a 0.8% rate of trial participation (see Appendix Table A6. By contrast, a patient at the 90th percentile of the risk distribution has SAE treatment effects that are over 2.5 times

larger (3.6 percentage points per month), yet has a 4 times smaller chance of trial participation (0.2%).

Next, we split our trial participation cohort into two subgroups: advanced stage cancers and early stage cancers. When constructing the target population for advanced stage cancers, we require that patients be receiving a cancer drug therapy that is specifically indicated for this disease stage.[9] This ensures that all patients in the comparator population are being actively treated—the sample will necessarily exclude patients for whom treatment is not consistent with their goals of care or for whom physicians judge that treatment would offer no clinical benefit due to the patient's frailty or disease state.

In the subsample with advanced stage cancers, the average trial participation rate is more than twice as high as in the full population, and the slope on the linear relationship between SAE risk and trial participation is approximately twice as large in magnitude, as reported in Table 5 column 2. Patients with 2.5 percentage points higher SAE risk (corresponding to the interquartile range of the SAE risk distribution in the trial participation cohort) are 0.6 percentage points less likely to participate in trials, relative to a mean trial participation rate of 1.1%.

Table 5 column 3 reports results for patients with early stage cancers; this cohort does not limit to patients who currently receive cancer drugs, because the standard of care for many of these early stage cancers do not require treatment with drugs. (Early stage cancers are commonly treated with surgical resection and/or radiation.) As a result, clinical trials targeting this population aim to introduce drug treatment to a population not currently receiving drugs. We cannot know for certain which patients among this early-stage cohort might receive a new drug, so we include all patients with the relevant cancer type, tumor markers, and stage in our analysis. We continue to find a strong, negative relationship between SAE risk and trial participation. The mean trial participation rates is 36% as large for early stage cancers relative to advanced stage cancers, and the slope of the relationship between SAE risk and trial participation is 35% as large as in the advanced stage sample. A 2.5 percentage point increase in the SAE rate implies a 0.2 percentage

[9] We do this in part because while our data identifies cancer stage at diagnosis, we cannot directly observe disease progression after diagnosis. Thus, we follow the advice of oncologist experts and use treatment using drugs that are specifically used for advanced cancers to identify patients with advanced disease stages.

point decline in trial participation rate, from a mean participation rate of 0.4%.

We also assess the relationship between SAE risk and trial participation conditional on patient demographics. The FDA has focused attention on improving trial generalizability by ensuring that trial enrollees match the age, sex, and race distribution of the target population (Food and Drug Administration Oncology Center of Excellence, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research, 2022; **?**). Would improving the demographic representativeness of trial enrollees be sufficient to close the estimated gaps in SAE effects? Table 5 Panel B reveals that controlling for sex, race, and age (with linear and quadratic terms) does little to change the estimated relationship between SAE risk and trial participation. The pooled sample coefficient attenuates only slightly with the addition of these control variables, from a baseline estimate of -0.128 to a controlled estimate of -0.101. This suggests that improving demographic representation of clinical trials is insufficient for ensuring that SAE treatment effects generalize. Most of the selection of lower-risk patients into trial enrollment occurs within demographic groups. Appendix Table A5 replicates this analysis using logistic regression (rather than a linear probability model) and reports consistent results.

5.3 Implications for Estimating SAE Treatment Effects in Trials

Finally, we consider the implications of our findings for the average SAE risk reported in trials. Figure 5 shows kernel density plots of the distribution of SAE risk among all patients in our cohort, and among trial participants. The distribution of SAE risk for trial enrollees has a lower mode and less variation compared to matched cohort of non-enrollees. The average SAE risk for trial participants is 3.1%, compared to a risk of 3.8% among all patients. Using the relationship between SAE risk and SAE treatment effects estimated in Section 4.2, this difference corresponds to an average SAE treatment effect of 2.2 percentage points for trial participants compared to an average SAE treatment effect of 2.5 percentage points for the full population, a 15% difference.

Our sample is limited to patients who met the age eligibility criterion for Medicare for at least 1 year prior to cancer diagnosis, and are thus at least 66 years old. Our findings suggest that subgroup analyses of patients over 65 are likely to understate true SAE treatment effects. The true

gap between SAE treatment effect for trial enrollees and the target population is likely greater, since patients over 65 years old are underrepresented in clinical trials (Varma et al., 2023). Because older patients are more likely to be frail and have multiple comorbidities, SAE treatment effects likely increase with age. Thus, underrepresentation of patients over 65 in trials would further widen the gap between trial participants and the target population.

In Appendix Figure A3, we further explore whether industry-sponsored trials have different patterns of selection compared to investigator-initiated trials. Consistent with the possibility that industry-sponsored trials have stronger incentives to reduce SAE among trial participants, we find that industry-sponsored trials are less likely to draw participants with above-average SAE risk. However, the difference in mean SAE risk across industry-sponsored vs. non-industry sponsored trials is modest: 3.1 vs. 3.3 percentage points per month. It is worth noting that even among trials initiated by an academic investigator, many report funding from pharmaceutical companies; thus, industry preferences may influence aspects of trial design for both types of trials analyzed here.

Finally, we find that clinical trials enrollment would need to be twice as large to conduct adequately powered subgroup analyses of SAE treatment effects in high-risk groups under current enrollment patterns compared to representative enrollment. We conducted power calculations to find the trial size needed to detect a treatment effect on SAE outcomes among a high-risk subgroup, with 80% power at the 5% significance level. The high-risk group is defined as the top 30% of SAE risk distribution in our trial participation cohort; monthly average SAE risk is 6.1% in this group, which is 60% larger than the overall mean. We assume a three-month trial duration and monthly SAE treatment effect as estimated in Table 4 Column 2. Under current trial enrollment patterns, only 15% of trial enrollees are in the top 30% of the risk distribution, and a trial would need to enroll 2273 total participants to have the 350 high-risk participants needed to power subgroup analyses. If enrollment were proportional to the population, then a trial would need to enroll half as many participants (1133) in order to have 350 high-risk participants needed to power subgroup analyses.

6 Regulating SAE Representativeness

So far, we have presented evidence that trials are not representative in terms of the SAE harms faced by patients likely to take up the tested drugs. We now propose a simple model of how representativeness in this dimension might be regulated and how trial sponsors could respond.

We propose a regulation consisting of five parts:

1. Trialists identify the target population should the drug be approved.

2. Trialists compute means of observable covariates in the target population and in the proposed trial population, denoted by $\bar{X}^{target}$ and $\bar{X}^{trial}$.

3. Given these means, use the LASSO model from Equation 1 and the scaling coefficient $\hat{\theta}_1$ from Equation 5 to compute $B_\gamma = \hat{\theta}_1(\bar{X}^{target} - \bar{X}^{trial})'\hat{\beta}$, the estimated SAE bias. Refine trial enrollment strategy to minimize predicted SAE bias prior to approval.

4. When submitting trial findings for FDA review, report the target population definition, predicted SAE risk in the target population, and predicted SAE risk among enrolled patients.

5. In postmarketing surveillance, confirm alignment between predicted target population and realized target population (subject to fines or other penalties if there is misalignment).

The above regulation references the specific SAE risk model we develop here, which would be appropriate for oncology drugs. Adapting the framework to other disease targets would require developing an appropriate SAE risk model for those populations. Our work on cancer drugs suggests a blueprint for this future extension to other settings.

The regulatory structure outlined above echoes aspects of the FDA's 2024 draft guidance for ensuring race, ethnicity, sex, and age group diversity in trials.[10] The FDA guidance proposed that trials should pre-specify target levels of enrollment for each demographic group, and justify those goals: "enrollment goals should be informed by the estimated prevalence or incidence of the disease or condition in the U.S. intended use population" (U.S. Food & Drug Administration, 2024). Our proposed regulation includes this structure of pre-specification and justification, while

[10]This draft guidance was subsequently withdrawn by the FDA in 2025.

focusing attention on a risk index that is likely to predict the drug's safety risks. Further, our approach provides a quantifiable metric of predicted bias given the trial's enrollment patterns, for the safety risks commonly associated with cancer drugs. We depart from the FDA's proposed regulatory structure by suggesting a need for postmarket surveillance to ensure that trial sponsors have incentives to accurately specify the target population. An advantage of our approach is that it provides a single, clinically-relevant index for reporting trial enrollment, overcoming barriers associated with simultaneously monitoring enrollment across many dimensions.

A key question for such a regulation is how trial sponsors would respond—specifically, will improving representativeness of severe adverse event risk make trial results more or less representative of net impacts? In Appendix A1, we formalize this idea by assuming that trial sponsors seeking unbiasedness in terms of γ (severe adverse events) will shift the underlying shares of enrollees from each subgroup in the way that achieves unbiasedness while changing the Euclidean distance between the shares as little as possible.[11] Define $B_\alpha = \sum_j (r_j - p_j)\alpha_j$ (the initial α-bias) and $B_\gamma = \sum_j (r_j - p_j)\gamma_j$ (the initial γ-bias). Define $h = \frac{Cov_p(\alpha,\gamma)}{Var_p(\gamma)}$, the population regression weighted coefficient of α on γ. In the case where B_α and B_γ have the same sign (or $B_\alpha = 0$), achieving γ unbiasedness will reduce τ unbiasedness when:

$$\max(-h, h - 2\frac{B_\alpha}{B_\gamma}) \leq \delta \tag{9}$$

Note that if $h = 0$, this always holds since $\delta > 0$. When harms γ_j and benefits α_j are positively correlated, reducing γ unbiasedness will also tend to reduce τ (net benefit) unbiasedness unless γ and α are sufficiently correlated that achieving unbiasedness in terms of γ leads to large increases in the α bias. (The second term inside the max highlights that if α changes in a direction that initially mitigates bias, α must "overshoot" by enough to generate substantial bias in the other direction.)

[11] Shifting shares in this way would require trial sponsors to know r_j and p_j before conducting trials. In the appendix, we assume that in trial k, net impacts are given by $\tau_{j,k} = \beta_k(\alpha_j + \delta\gamma_j)$. Trialists can then know $E(\beta_k)\alpha_j$ and $E(\beta_k)\gamma_j$ for each subgroup from pooling evidence from previous trials (as we do in our empirical work above) while still requiring a trial to estimate τ_j because of the β_k term.

Model Calibration We conduct a calibration exercise to assess whether the condition in Equation 9 is likely to hold in practice, in the context of cancer drug trials. Details of this exercise are available in Appendix A3. We summarize the exercise here. To estimate h, the relationship between heterogeneous harms and benefits of treatment, we first estimate medical benefits by combining estimates of cancer mortality risk with estimates of the average mortality risk reduction from treatment (20%). Next, using these measures of treatment benefits and harms, we relate α and γ. We find that $h = -0.049$, suggesting only a weak relationship between SAE treatment effects and treatment benefits, whereby patients with a 10 percentage point (p.p.) greater probability of a drug-induced SAE event would experience a 0.5 p.p. greater decline in 1-year cancer mortality from treatment.

In our empirical work above, we presented evidence that B_γ is negative, i.e., trials understate the risk of drug-induced harms. We do not have direct evidence about the sign of B_α, but we suspect it is likely negative or zero if trials select patients who are likely to have greater benefits (consistent with trial sponsors' incentives to demonstrate that the drug is effective). Given that $h < 0$, if $B_\alpha \leq 0$, then comparing $-h = 0.049$ to δ is sufficient for checking the condition in Equation 9. The δ parameter has only been previously estimated given hypothetical preference elicitation surveys, but we find that the existing estimates of the utility cost of an SAE are roughly 4 times larger than would be required for the condition in equation 9 to hold, meaning that regulating SAE representativeness alone is likely to improve net benefit representativeness for cancer drugs.

7 Conclusion

This paper finds that cancer patients who are at risk of serious adverse events are substantially less likely to be enrolled in clinical trials yet more likely to experience adverse events as a causal result of drug treatment. There are several possible reasons for systematic underrepresentation of high SAE risk and high SAE treatment effect patients in trials. Pharmaceutical companies themselves may not want to enroll patients who are likely to develop complications. In addition, physicians enrolling patients in trials may be reluctant to enroll patients who would require additional monitoring or care. Patient preferences may also play a role; Alsan et al. (2024) finds that a history of

underrepresentation may make patients reluctant to enroll, if they anticipate smaller benefits of participation. Our results cannot separate these channels, although anecdotally, a pharmaceutical executive (who might be incentivized to blame other parties) privately emphasized the first channel as paramount.

Assuming external validity of trials' net impacts is a desirable goal, our results imply that current trials depart from this normative standard in systematic ways that regulators could monitor, provided they collected data on average rates of comorbidities in trials and target clinical populations. Our results do not speak to a broader question about the desirability of representativeness. If some types of patients are more costly to enroll than others, then there is a trade-off between representativeness and convenience sampling. Additionally, there are trade-offs between representativeness and statistical power: for example, it may be desirable to enroll patients at higher risk of the treated condition to maximize power to detect proportional treatment effects, at the expense of externally valid estimates of adverse events. These important issues are beyond the scope of our investigation here.

References

Agha, Leila, Douglas Staiger, Christopher Brown, Roger F. Soll, Jeffrey D. Horbar, and Erika M. Edwards, "Association of hospital adoption of probiotics with outcomes among neonates with very low birth weight," *JAMA Health Forum*, May 2023, *4* (5), e230960.

_ , Soomi Kim, and Danielle Li, "Insurance design and pharmaceutical innovation," *American Economic Review: Insights*, 2022, *4* (2), 191–208.

Al-Ubaydli, Omar, John A List, and Dana L Suskind, "What can we learn from experiments? Understanding the threats to the scalability of experimental results," *American Economic Review*, 2017, *107* (5), 282–286.

Allcott, Hunt, "Site selection bias in program evaluation," *The Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

Alosh, Mohamed, Kathleen Fritsch, Mohammad Huque, Kooros Mahjoob, Gene Pennello, Mark Rothmann, Estelle Russek-Cohen, Fraser Smith, Stephen Wilson, and Lilly Yue, "Statistical considerations on subgroup analysis in clinical trials," *Statistics in Biopharmaceutical Research*, 2015, *7* (4), 286–303.

Alsan, Marcella, Maya Durvasula, Harsh Gupta, Joshua Schwartzstein, and Heidi Williams, "Representation and extrapolation: evidence from clinical trials," *The Quarterly Journal of Economics*, 2024, *139* (1), 575–635.

Azoulay, Pierre, "Do pharmaceutical sales respond to scientific evidence?," *Journal of Economics & Management Strategy*, 2002, *11* (4), 551–594.

Barsky, Robert, John Bound, Kerwin Ko' Charles, and Joseph P Lupton, "Accounting for the black–white wealth gap: a nonparametric approach," *Journal of the American Statistical Association*, 2002, *97* (459), 663–673.

Beusterien, KM, SM Szabo, S Kotapati, J Mukherjee, A Hoos, P Hersey, MR Middleton, and AR Levy, "Societal preference values for advanced melanoma health states in the United Kingdom and Australia," *British Journal of Cancer*, 2009, *101* (3), 387–389.

Bishnoi, Rohit, Zhigang Xie, Chintan Shah, Jiang Bian, Hemant S. Murthy, John R. Wingard, and Nosha Farhadfar, "Real-world experience of carfilzomib-associated cardiovascular adverse events: SEER-Medicare data set analysis," *Cancer Medicine*, January 2021, *10* (1), 70–78.

Bittoni, Marisa A., Ashwini Arunachalam, Haojie Li, Ramon Camacho, Jinghua He, Yichen Zhong, Gregory M. Lubiniecki, and David P. Carbone, "Real-World Treatment Patterns, Overall Survival, and Occurrence and Costs of Adverse Events Associated With First-line Therapies for Medicare Patients 65 Years and Older With Advanced Non-small-cell Lung Cancer: A Retrospective Study," *Clinical Lung Cancer*, September 2018, *19* (5), e629–e645.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event-study designs: robust and efficient estimation," *Review of Economic Studies*, 2024, *91* (6), 3253–3285.

Bradley, Cathy J., Charles W. Given, Bassam Dahman, and Timothy L. Fitzgerald, "Adjuvant chemotherapy after resection in elderly Medicare and Medicaid patients with colon cancer," *Archives of Internal Medicine*, March 2008, *168* (5), 521–529.

Budish, Eric, Benjamin N Roin, and Heidi Williams, "Do firms underinvest in long-term research? Evidence from cancer clinical trials," *American Economic Review*, 2015, *105* (7), 2044–2085.

Busch, Susan H, Ezra Golberstein, and Ellen Meara, "The FDA and ABCs: unintended consequences of antidepressant warnings on human capital," *Journal of Human Resources*, 2014, *49* (3), 540–571.

Chandra, Amitabh, Pragya Kakani, and Adam Sacarny, "Hospital allocation and racial disparities in health care," *Review of Economics and Statistics*, 2024, *106* (4), 924–937.

Chassang, Sylvain and Samuel Kapon, "Designing randomized controlled trials with external validity in mind," Technical Report, National Bureau of Economic Research 2022.

Chintagunta, Pradeep K, Renna Jiang, and Ginger Z Jin, "Information, learning, and drug diffusion: The case of Cox-2 inhibitors," *QME*, 2009, *7* (4), 399–443.

Cochran, William G., *Sampling Techniques*, 3rd ed., New York: John Wiley & Sons, 1977.

Cohen, Lauren, Umit G Gurun, and Danielle Li, "Internal deadlines, drug approvals, and safety problems," *American Economic Review: Insights*, 2021, *3* (1), 67–82.

Connolly, Anna, Marcia Kirwan, and Anne Matthews, "A scoping review of the methodological approaches used in retrospective chart reviews to validate adverse event rates in administrative data," *International Journal for Quality in Health Care*, 2024, *36* (2), mzae037.

Dahabreh, Issa J, Sarah E Robertson, Jon A Steingrimsson, Elizabeth A Stuart, and Miguel A Hernan, "Extending inferences from a randomized trial to a new target population," *Statistics in Medicine*, 2020, *39* (14), 1999–2014.

_ , Sebastien JP A Haneuse, James M Robins, Sarah E Robertson, Ashley L Buchanan, Elizabeth A Stuart, and Miguel A Hernán, "Study designs for extending causal inferences from a randomized trial to a target population," *American Journal of Epidemiology*, 2021, *190* (8), 1632–1642.

Degtiar, Irina and Sherri Rose, "A review of generalizability and transportability," *Annual Review of Statistics and Its Application*, 2023, *10* (1), 501–524.

Downing, Nicholas S., Nilay D. Shah, Jenerius A. Aminawung, Alison M. Pease, Jean-David Zeitoun, Harlan M. Krumholz, and Joseph S. Ross, "Postmarket Safety Events Among Novel Therapeutics Approved by the US Food and Drug Administration Between 2001 and 2010," *JAMA*, May 2017, *317* (18), 1854–1863.

Du, Xianglin L., Wenyaw Chan, Sharon Giordano, Jane M. Geraci, George L. Delclos, Keith Burau, and Shenying Fang, "Variation in modes of chemotherapy administration for

breast carcinoma and association with hospitalization for chemotherapy-related toxicity," *Cancer*, September 2005, *104* (5), 913–924.

Dy, Sydney M., Phoebe Sharkey, Robert Herbert, Kathleen Haddad, and Albert W. Wu, "Comorbid illnesses and health care utilization among Medicare beneficiaries with lung cancer," *Critical Reviews in Oncology/Hematology*, September 2006, *59* (3), 218–225.

Food and Drug Administration Oncology Center of Excellence, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research, "Inclusion of Older Adults in Cancer Clinical Trials," March 2022.

Freedman, Rachel A., Ines Vaz-Luis, William T. Barry, Huichuan Lii, Nancy U. Lin, Eric P. Winer, and Nancy L. Keating, "Patterns of chemotherapy, toxicity, and short-term outcomes for older women receiving adjuvant trastuzumab-based therapy," *Breast Cancer Research and Treatment*, June 2014, *145* (2), 491–501.

_ , Katherine S. Virgo, Yulei He, Alexandre L. Pavluck, Eric P. Winer, Elizabeth M. Ward, and Nancy L. Keating, "The association of race/ethnicity, insurance status, and socioeconomic factors with breast cancer care," *Cancer*, 2011, *117* (1), 180–189. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cncr.25542.

Fujimoto, Daichi, Takeshi Morimoto, Motohiro Tamiya, Akito Hata, Hirotaka Matsumoto, Atsushi Nakamura, Toshihide Yokoyama, Yoshihiko Taniguchi, Junji Uchida, Yuki Sato, Takashi Yokoi, Hisashi Tanaka, Naoki Furuya, Takeshi Masuda, Yoshihiko Sakata, Eisaku Miyauchi, Satoshi Hara, Go Saito, Satoru Miura, Masaki Kanazu, Nobuyuki Yamamoto, and Hiroaki Akamatsu, "Outcomes of Chemoimmunotherapy Among Patients With Extensive-Stage Small Cell Lung Cancer According to Potential Clinical Trial Eligibility," *JAMA Network Open*, Feb 2023, *6* (2), e230698.

Gardner, John, "Two-stage difference-in-differences," *arXiv preprint arXiv:2207.05943v1*, 2022.

George, Mathew, Alexandra Smith, Sabe Sabesan, and Geetha Ranmuthugala, "Physical Comorbidities and Their Relationship with Cancer Treatment and Its Outcomes in Older Adult Populations: Systematic Review," *JMIR cancer*, October 2021, *7* (4), e26425.

Green, Angela K., Sara M. Tabatabai, Carol Aghajanian, Ola Landgren, Gregory J. Riely, Paul Sabbatini, Peter B. Bach, Colin B. Begg, Allison Lipitz-Snyderman, and Sham Mailankody, "Clinical trial participation among older adult Medicare Fee-for-Service beneficiaries with cancer," *JAMA Oncology*, December 2022, *8* (12), 1786–1792.

_ , _ , Xing Bai, Akriti Mishra Meza, Anne-Marie Lesny, Carol Aghajanian, Ola Landgren, Gregory J. Riely, Paul Sabbatini, Andrew Salner, Scott Lipkin, Andrew Ip, Peter B. Bach, Colin B. Begg, Sham Mailankody, and Allison Lipitz-Snyderman, "Validation of a population-based data source to examine national cancer clinical trial participation," *JAMA Network Open*, March 2022, *5* (3), e223687.

Grønberg, Bjørn H., Stein Sundstrøm, Stein Kaasa, Roy M. Bremnes, Oystein Fløtten, Tore Amundsen, Harald H. Hjelde, Christian von Plessen, and Marit Jordhøy, "Influence of comorbidity on survival, toxicity and health-related quality of life in patients with advanced non-small-cell lung cancer receiving platinum-doublet chemotherapy," *European Journal of Cancer (Oxford, England: 1990)*, August 2010, *46* (12), 2225–2234.

Gunturu, Krishna S., Timothy T. Pham, Sonali Shambhu, Michael J. Fisch, John J. Barron, and David Debono, "Immune checkpoint inhibitors: immune-related adverse events, healthcare utilization, and costs among commercial and Medicare Advantage patients," *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer*, May 2022, *30* (5), 4019–4026.

Hanlon, Peter, Elaine Butterly, Anoop SV Shah, Laurie J Hannigan, Sarah H Wild, Bruce Guthrie, Frances S Mair, Sofia Dias, Nicky J Welton, and David A McAllister, "Assessing trial representativeness using serious adverse events: an observational analysis using aggregate and individual-level data from clinical trials and routine healthcare data," *BMC medicine*, 2022, *20* (1), 410.

Hansen, Ryan N., Scott D. Ramsey, Deepa Lalla, Anthony Masaquel, Tripthi Kamath, Melissa Brammer, Sara A. Hurvitz, and Sean D. Sullivan, "Identification and cost of adverse events in metastatic breast cancer in taxane and capecitabine based regimens," *SpringerPlus*, 2014, *3*, 259.

Hassett, Michael J., Sowmya R. Rao, Suzana Brozovic, James E. Stahl, Joel H. Schwartz, Betty Maloney, and Joseph O. Jacobson, "Chemotherapy-related hospitalization among community cancer center patients," *The Oncologist*, 2011, *16* (3), 378–387.

Herbach, Emma, Michael A. O'Rorke, Ryan M. Carnahan, Bradley D. McDowell, Bryan Allen, Isabella Grumbach, Barry London, Brian J. Smith, Douglas R. Spitz, Aaron Seaman, and Elizabeth A. Chrischilles, "Cardiac Adverse Events Associated With Chemo-Radiation Versus Chemotherapy for Resectable Stage III Non-Small-Cell Lung Cancer: A Surveillance, Epidemiology and End Results-Medicare Study," *Journal of the American Heart Association*, December 2022, *11* (23), e027288.

Hershman, Dawn L., Jason D. Wright, Emerson Lim, Donna L. Buono, Wei Yann Tsai, and Alfred I. Neugut, "Contraindicated use of bevacizumab and toxicity in elderly patients with cancer," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, October 2013, *31* (28), 3592–3599.

Hughes, Shannon, David Cohen, and Rachel Jaggi, "Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: a cross-sectional study," *BMJ open*, 2014, *4* (7), e005535.

Hurria, Arti, Kayo Togawa, Supriya G. Mohile, Cynthia Owusu, Heidi D. Klepin, Cary P. Gross, Stuart M. Lichtman, Ajeet Gajra, Smita Bhatia, Vani Katheria, Shira Klapper, Kurt Hansen, Rupal Ramani, Mark Lachs, F. Lennie Wong, and William P. Tew, "Predicting chemotherapy toxicity in older adults with cancer: a prospective multicenter study," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, September 2011, *29* (25), 3457–3465.

Jain, P., J. Gutierrez Bugarin, A. Guha, C. Jain, N. Patil, T. Shen, I. Stanevich, V. Nikore, K. Margolin, M. Ernstoff, V. Velcheti, J. Barnholtz-Sloan, and A. Dowlati, "Cardiovascular adverse events are associated with usage of immune checkpoint inhibitors in real-world clinical data across the United States," *ESMO Open*, Oct 2021, *6* (5), 100252.

Kao, Jennifer L and Tamar Oostrom, "Representation in product development: Evidence from insurance and clinical trials," *Available at SSRN 4928902*, 2024.

Kennedy-Martin, Tessa, Sarah Curtis, Douglas Faries, Susan Robinson, and Joseph Johnston, "A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results," *Trials*, 2015, *16*, 1–14.

Kim, Dae Hyun, Sebastian Schneeweiss, Robert J Glynn, Lewis A Lipsitz, Kenneth Rockwood, and Jerry Avorn, "Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index," *The Journals of Gerontology: Series A*, June 2018, *73* (7), 980–987.

Lackey, Leila, Graham Thompson, and Sara Eggers, "FDA's benefit–risk framework for human drugs and biologics: role in benefit–risk assessment and analysis of use for drug approvals," *Therapeutic innovation & regulatory science*, 2021, *55* (1), 170–179.

Luo, RuiLi, Sharon H. Giordano, Jean L. Freeman, Dong Zhang, and James S. Goodwin, "Referral to medical oncology: a crucial step in the treatment of older patients with stage III colon cancer," *The Oncologist*, October 2006, *11* (9), 1025–1033.

Manski, Charles F, "Improving clinical guidelines and decisions under uncertainty," Technical Report, National Bureau of Economic Research 2017.

McKibbin, Rebecca, "The effect of RCTs on drug demand: Evidence from off-label cancer drugs," *Journal of Health Economics*, 2023, *90*, 102779.

Michelman, Valerie and Lucy Msall, "Sex, Drugs, and RD: Missing Innovation from Regulating Female Enrollment in Clinical Trials," 2024.

Muss, Hyman B., Donald A. Berry, Constance Cirrincione, Daniel R. Budman, I. Craig Henderson, Marc L. Citron, Larry Norton, Eric P. Winer, Clifford A. Hudis, and Cancer and Leukemia Group B Experience, "Toxicity of older and younger patients treated with adjuvant chemotherapy for node-positive breast cancer: the Cancer and Leukemia Group B Experience," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, August 2007, *25* (24), 3699–3704.

Nadkarni, Prakash M, "Drug safety surveillance using de-identified EMR and claims data: issues and challenges," *Journal of the American Medical Informatics Association*, 2010, *17* (6), 671–674.

National Cancer Institute, "Age and Cancer Risk," May 2025.

Neugut, Alfred I., Matthew Matasar, Xiaoyan Wang, Russell McBride, Judith S. Jacobson, Wei-Yann Tsai, Victor R. Grann, and Dawn L. Hershman, "Duration of adjuvant chemotherapy for colon cancer and survival among the elderly," *Journal of Clinical Oncology*, May 2006, *24* (15), 2368–2375. Publisher: Wolters Kluwer.

Neyman, Jerzy, "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," *Journal of the Royal Statistical Society. Series A (General)*, 1934, *97* (4), 558–625.

Nordon, Clementine, Helene Karcher, Rolf HH Groenwold, Mikkel Zöllner Ankarfeldt, Franz Pichler, Helene Chevrou-Severac, Michel Rossignol, Adeline Abbe, Lucien Abenhaim et al., "The "efficacy-effectiveness gap": historical background and current conceptualization," *Value in Health*, 2016, *19* (1), 75–81.

of Sciences, Engineering National Academies and Medicine, *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*, Washington, DC: The National Academies Press, 2022.

O'Grady, Margaret A., Elyse Slater, Elin R. Sigurdson, Neal J. Meropol, Alan Weinstein, Charles J. Lusch, Elaine Sein, Patricia Keeley, Bonnie Miller, Paul F. Engstrom, and Steven J. Cohen, "Assessing compliance with national comprehensive cancer

network guidelines for elderly patients with stage III colon cancer: the Fox Chase Cancer Center Partners' initiative," *Clinical Colorectal Cancer*, June 2011, *10* (2), 113–116.

Olson, Mary K, "Pharmaceutical policy change and the safety of new drugs," *The Journal of Law and Economics*, 2002, *45* (S2), 615–642.

_ , "The risk we bear: the effects of review speed and industry user fees on new drug safety," *Journal of health economics*, 2008, *27* (2), 175–200.

Oostrom, Tamar, "Funding of clinical trials and reported drug efficacy," *Journal of Political Economy*, 2024, *132* (10), 3298–3333.

Orcutt, Xavier, Kan Chen, Ronac Mamtani, Qi Long, Ravi B. Parikh et al., "Evaluating generalizability of oncology trial results to real-world patients using machine learning-based trial emulations," *Nature Medicine*, 2025, *31*, 457–465.

Pasetto, Lara Maria, Cristina Falci, Umberto Basso, Giampietro Gasparini, Mario D'Andrea, Paola Bonginelli, Emilio Bajetta, Marco Platania, Oscar Alabiso, Stefania Miraglia, Erica Bertona, Francesco Oniga, Rita Biason, Maria Concetta Chetrì, Palma Fedele, Giovanna Massara, Incoronata Romaniello, Maria Emanuela Negru, Giovanna Luchena, Monica Giordano, Franco Buzzi, Riccardo Ricottao, Salvatore Sienao, and Silvio Monfardini, "Adjuvant treatment for elderly patients with colon cancer. An observational study," *Anticancer Research*, 2008, *28* (4C), 2513–2518.

Philipson, Tomas, Ernst R Berndt, Adrian HB Gottschalk, and Eric Sun, "Cost-benefit analysis of the FDA: The case of the prescription drug user fee acts," *Journal of Public Economics*, 2008, *92* (5-6), 1306–1325.

Philipson, Tomas J and Eric Sun, "Is the Food and Drug Administration safe and effective?," *Journal of Economic Perspectives*, 2008, *22* (1), 85–102.

Rashid, Nazia, Han A. Koh, Hilda C. Baca, Zhaoliang Li, Susan Malecha, Oyewale Abidoye, and Anthony Masaquel, "Clinical Impact of Chemotherapy-Related Adverse

Events in Patients with Metastatic Breast Cancer in an Integrated Health Care System," *Journal of Managed Care & Specialty Pharmacy*, October 2015, *21* (10), 863–871.

Reeder-Hayes, Katherine E., Anne Marie Meyer, Sharon Peacock Hinton, Ke Meng, Lisa A. Carey, and Stacie B. Dusetzina, "Comparative Toxicity and Effectiveness of Trastuzumab-Based Chemotherapy Regimens in Older Women With Early-Stage Breast Cancer," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, October 2017, *35* (29), 3298–3305.

Rivera-Caravaca, José Miguel, María Asunción Esteve-Pastor, Francisco Marín, Mariano Valdés, Vicente Vicente, Vanessa Roldán, and Gregory Y. H. Lip, "A Propensity Score Matched Comparison of Clinical Outcomes in Atrial Fibrillation Patients Taking Vitamin K Antagonists: Comparing the "Real-World" vs Clinical Trials," *Mayo Clin Proc*, Aug 2018, *93* (8), 1065–1073.

Robertson, Sarah E, Nina R Joyce, Jon A Steingrimsson, Elizabeth A Stuart, Denise R Aberle, Constantine A Gatsonis, and Issa J Dahabreh, "Comparing lung cancer screening strategies in a nationally representative US population using transportability methods for the National Lung Cancer Screening Trial," *JAMA Network Open*, 2024, *7* (1), e2346295–e2346295.

Sanoff, Hanna K., William R. Carpenter, Janet Freburger, Ling Li, Kun Chen, Leah L. Zullig, Richard M. Goldberg, Maria J. Schymura, and Deborah Schrag, "Comparison of adverse events during 5-fluorouracil versus 5-fluorouracil/oxaliplatin adjuvant chemotherapy for stage III colon cancer: a population-based analysis," *Cancer*, September 2012, *118* (17), 4309–4320.

Sarfati, Diana, Bogda Koczwara, and Christopher Jackson, "The impact of comorbidity on cancer and its treatment," *CA: A Cancer Journal for Clinicians*, July 2016, *66* (4), 337–350.

Shabaruddin, Fatiha H, Li-Chia Chen, Rachel A Elliott, and Katherine Payne, "A systematic review of utility values for chemotherapy-related adverse events," *Pharmacoeconomics*, 2013, *31* (4), 277–288.

Shah, Anuj, Ruchit Shah, Nehemiah Kebede, Ateesha Mohamed, Marc Botteman, Reg Waldeck, and Arif Hussain, "Real-world incidence and burden of adverse events among non-metastatic prostate cancer patients treated with secondary hormonal therapies following androgen deprivation therapy," *Journal of Medical Economics*, April 2020, *23* (4), 330–346.

Shah, Sachin J, Christopher R Manz, Brendan Balthis, Hari S Raman, Jason Abaluck, Nancy L Keating, and Leila Agha, "The role of regional and practice trial sites in nonrepresentative randomized cancer trial enrollment," *JNCI: Journal of the National Cancer Institute*, 2025, p. djaf071.

Sivendran, Shanthi, Asma Latif, Russell B McBride, Kristian D Stensland, Juan Wisnivesky, Lindsay Haines, William K Oh, and Matthew D Galsky, "Adverse event reporting in cancer clinical trial publications," *Journal of Clinical Oncology*, 2014, *32* (2), 83–89.

Steg, Philippe Gabriel, José López-Sendón, Esteban Lopez de Sa, Shaun G Goodman, Joel M Gore, Frederick A Anderson, Dominique Himbert, Jeanna Allegrone, Frans Van de Werf, Grace Investigators et al., "External validity of clinical trials in acute myocardial infarction," *Archives of Internal Medicine*, 2007, *167* (1), 68–73.

Tang, Eve, Philippe Ravaud, Carolina Riveros, Elodie Perrodeau, and Agnes Dechartres, "Comparison of serious adverse events posted at ClinicalTrials. gov and published in corresponding journal articles," *BMC medicine*, 2015, *13*, 1–8.

Tracy, Don, "New FDA Initiative Reveals Common Reasons for Drug Application Rejection," *Applied Clinical Trials*, July 11 2025.

U.S. Food & Drug Administration, "Enhancing the Diversity of Clinical Trial Populations-Eligibility Criteria, Enrollment Practices, and Trial Designs," November 2020.

_ , "Diversity Action Plans to Improve Enrollment of Participants from Underrepresented Populations in Clinical Studies," June 2024.

Varma, Tanvee, Michelle Mello, Joseph S. Ross, Cary Gross, and Jennifer Miller, "Metrics, baseline scores, and a tool to improve sponsor performance on clinical trial diversity: retrospective cross sectional study," *BMJ Medicine*, January 2023, *2* (1).

Vinod, Shalini K., Mark A. Sidhom, Gabriel S. Gabriel, Mark T. Lee, and Geoff P. Delaney, "Why do some lung cancer patients receive no anticancer treatment?," *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, July 2010, *5* (7), 1025–1032.

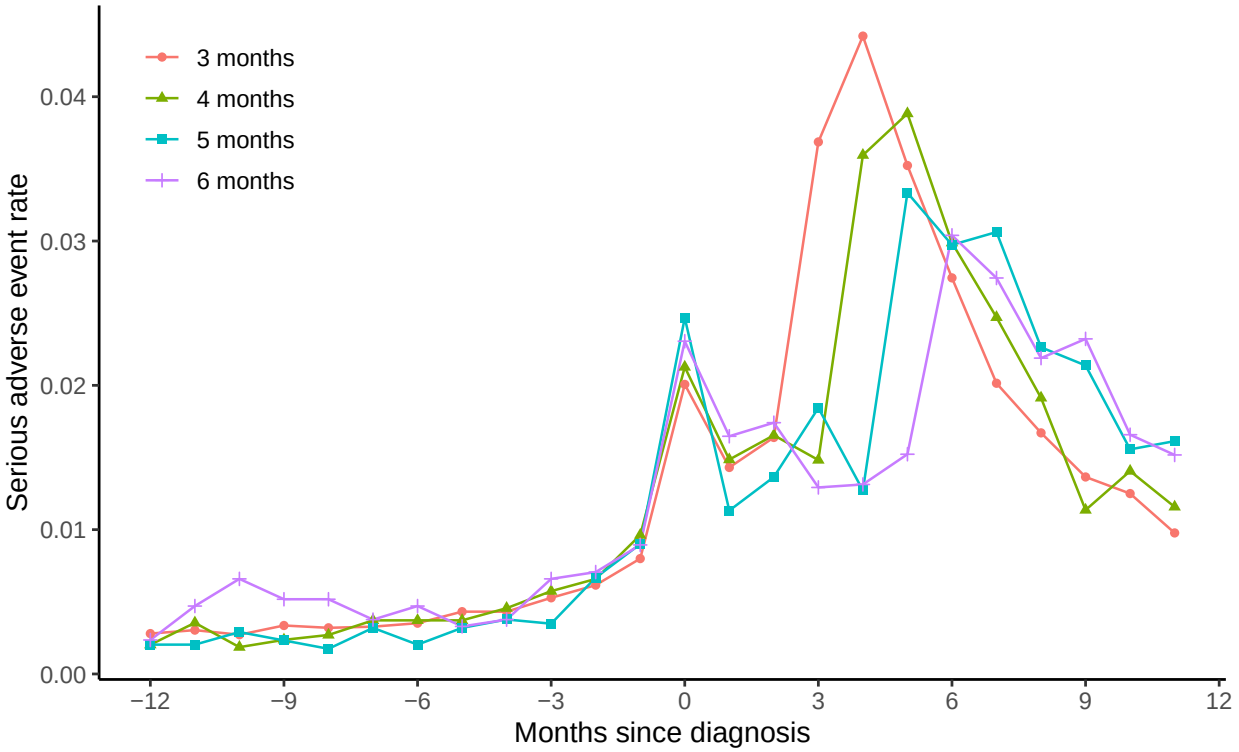
Visram, Alissa, Kelvin KW Chan, Hsien Seow, Gregory Pond, Ana Gayowsky, Arleigh McCurdy, Irwindeep Sandhu, Christopher P Venner, Guido Lancman, Amaris Balitsky et al., "Comparison of the efficacy in clinical trials versus effectiveness in the real-world of treatments for multiple myeloma: a population-based cohort study," *Blood*, 2023, *142*, 541.

Vivalt, Eva, "How much can we generalize from impact evaluations?," *Journal of the European Economic Association*, 2020, *18* (6), 3045–3089.

Wieder, Robert and Nabil Adam, "Racial Disparities in Breast Cancer Treatments and Adverse Events in the SEER-Medicare Data," *Cancers*, August 2023, *15* (17), 4333.

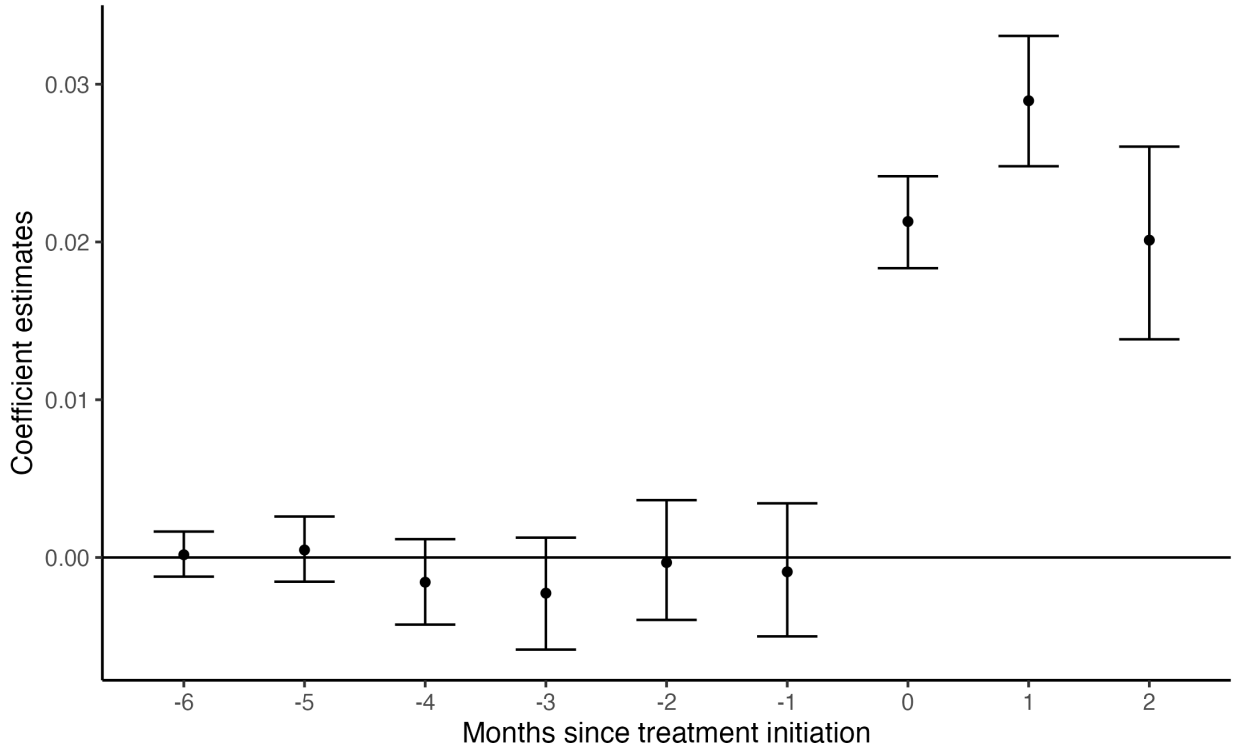
Zauderer, Marjorie, Sujata Patil, and Arti Hurria, "Feasibility and toxicity of dose-dense adjuvant chemotherapy in older women with breast cancer," *Breast Cancer Research and Treatment*, September 2009, *117* (1), 205–210.

Figure 1: Trends in Serious Adverse Event Rates, Before & After Diagnosis



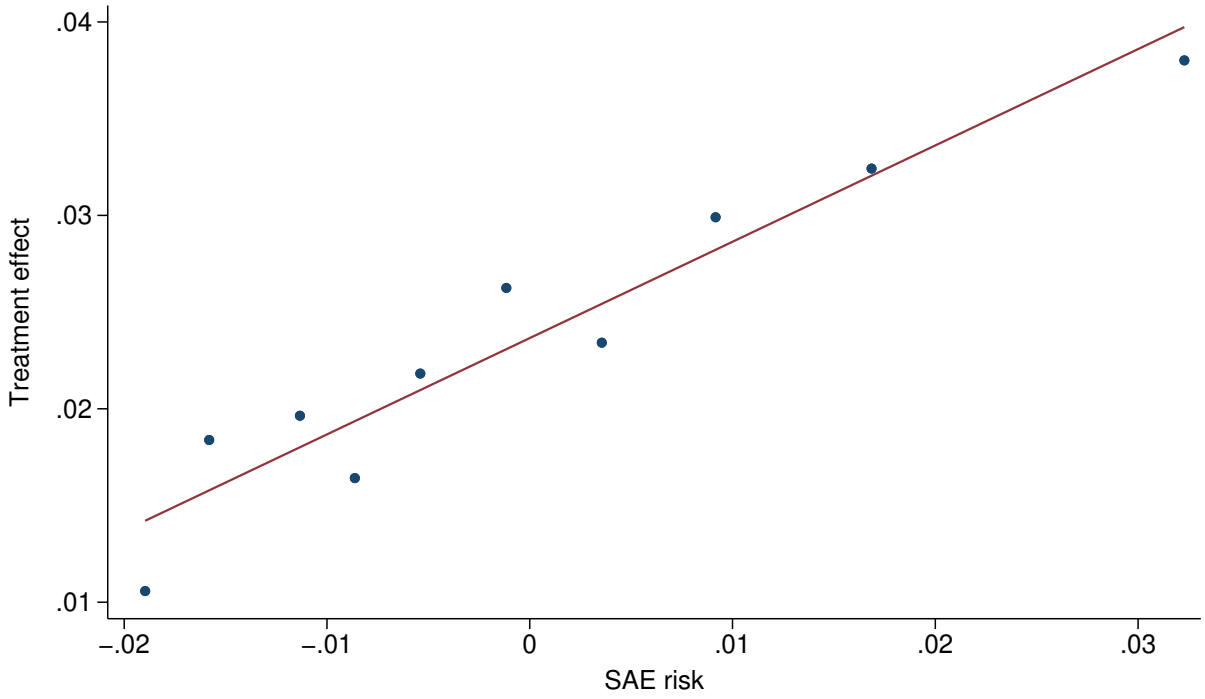
Notes: This figure shows the average SAE rate by month relative to diagnosis. The sample is split into 4 separate cohorts, corresponding to different timing of treatment initiation relative to diagnosis. The mean monthly SAE risk in the pre-diagnosis period is 0.0071. The SAE sample includes 24,002 patients.

Figure 2: Event Study of Serious Adverse Event Rate, Before & After Treatment Initiation



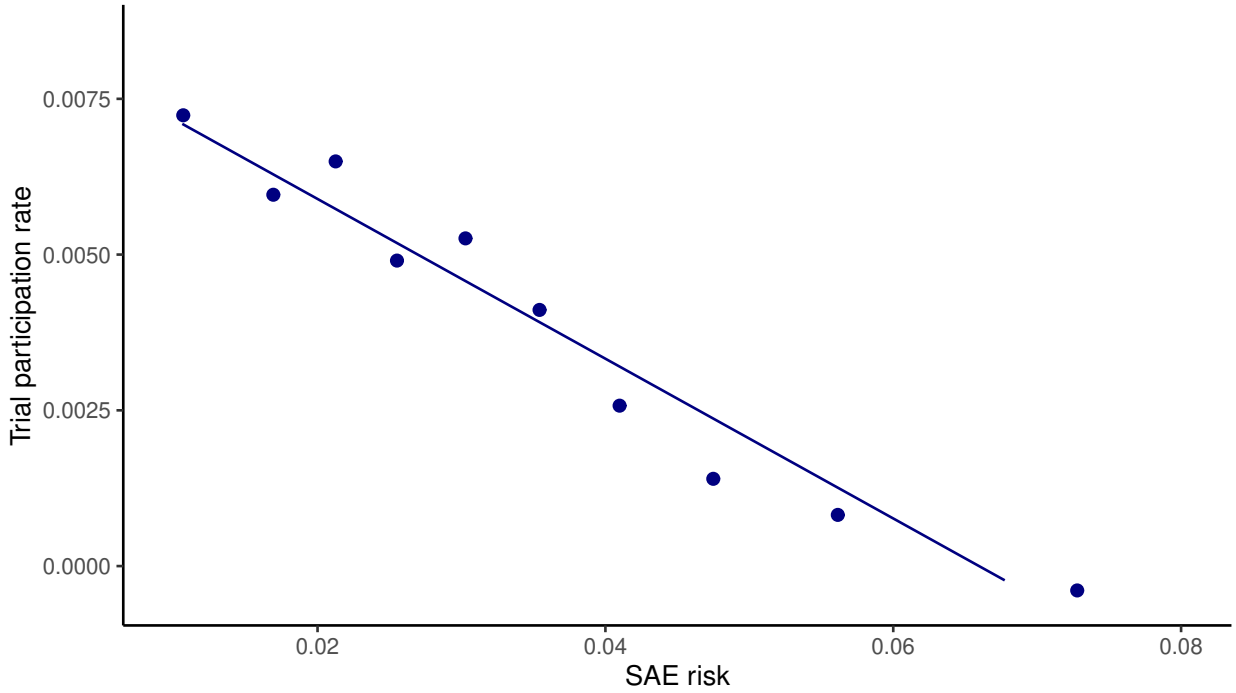
Notes: This figure displays coefficient estimates and bootstrapped 95% confidence intervals, for each month relative to treatment initiation. Pre-period coefficients come from estimating Equation 4; post-period values are averages of imputed treatment effects from the two-stage estimator. The outcome variable is a monthly indicator for any SAE. Control variables include fixed effects for: month since diagnosis, calendar time, treatment group (based on timing of treatment initiation 3, 4, 5, or 6 months after diagnosis), and cancer type by stage. We also control for patient age and predicted SAE risk. Confidence intervals are calculated using a bootstrap procedure, clustered at the patient level. The SAE sample includes 24,002 patients.

Figure 3: Relationship between Treatment Effect & Predicted Risk of Serious Adverse Events



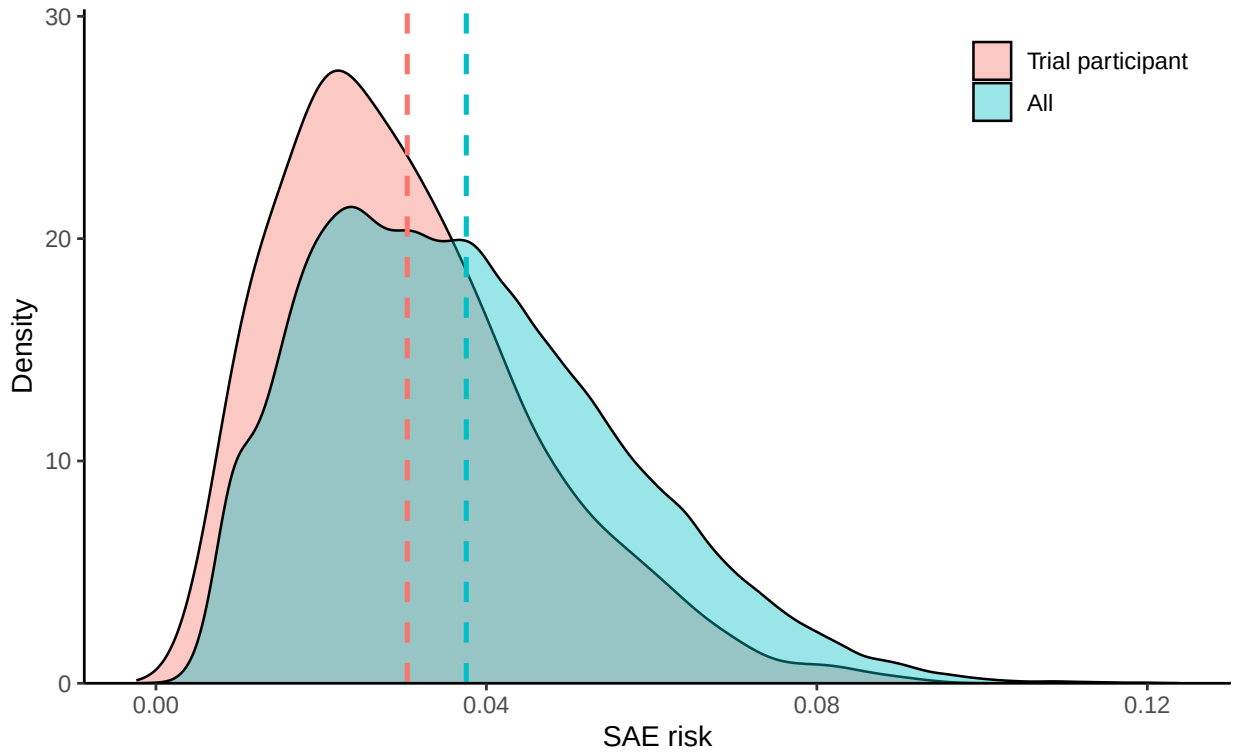
Notes: This binned scatterplot displays the relationship between individual-specific treatment effect estimates $\hat{\theta}$ from Equation 4 (calculated for treated individuals) and predicted SAE risk. Predicted SAE risk is measured in a sample that holds out the index observation. The SAE sample includes 24,002 patients.

Figure 4: Trial Participation & Predicted Risk of Serious Adverse Events



Notes: This binned scatterplot displays the relationship between trial participation and SAE risk in the full trial-eligible cohort, where SAE risk has been centered at the mean risk among the adverse event cohort. The binned scatterplot controls for disease state, defined as the interaction of cancer type, stage at diagnosis, treatment line, and biomarkers. The Trial Participation sample includes 157,926 patients.

Figure 5: Distribution of Predicted Serious Adverse Event Risk Among Trial Participants vs. Full Cohort



Notes: This figure displays kernel density plots of the SAE risk distribution among trial participants vs. the full trial-participation sample. Patients are re-weighted to match the distribution of specific disease cohorts for trial participants; disease cohorts are defined by the interaction of defined as the interaction of cancer type, stage at diagnosis, treatment line, and biomarkers (see Appendix Table A2). Mean risk equals 0.0309 for trial participants and 0.0380 in the full cohort. The Trial Participation sample includes 157,926 patients.

Table 1: Summary Statistics

	Adverse event cohort	Trial participation cohort	
	Treated in mo. 3-6	Non-participants	Trial participants
Age	74	75	72
Female	0.66	0.75	0.63
White	0.84	0.88	0.93
Black	0.09	0.07	0.04
Hypertension	0.74	0.81	0.73
Hyperlipidemia	0.74	0.81	0.74
Rheumatoid/osteoarthritis	0.50	0.58	0.51
Anemia	0.49	0.55	0.40
Ischemic heart disease	0.43	0.46	0.36
Cancer Types			
Bladder	X		
Breast	X	X	X
Colon	X		
Lung	X	X	X
Pancreas	X	X	X
Rectal	X		
Renal	X	X	X
Number of patients	24,002	157,074	852
Number of distinct trials			112

Notes: This table presents mean value of patient characteristics and summarizes the cancer types represented in the SAE and Trial Participation Samples.

Table 2: Differences between Trial Participants and Non-participants

	Non-participants	Trial participants	Absolute difference	Adjusted difference	
Female	0.75	0.63	-0.13	-0.01	
Kim frailty index	0.18	0.16	-0.02	-0.02	***
White	0.88	0.93	0.04	0.05	**
Black	0.07	0.04	-0.03	-0.03	**
Asian and Pacific Islander	0.04	0.03	-0.01	-0.01	
Age	74.6	71.7	-2.9	-2.7	***
Cataract	0.66	0.56	-0.11	-0.08	***
Diabetes	0.4	0.33	-0.07	-0.09	***
Ischemic heart disease	0.46	0.36	-0.09	-0.12	***
Depression	0.32	0.26	-0.05	-0.04	
Alzheimer's/dementia	0.1	0.04	-0.06	-0.05	***
Acute myocardial infarction	0.04	0.02	-0.01	-0.02	
Anemia	0.55	0.4	-0.15	-0.18	***
Asthma	0.16	0.12	-0.04	-0.04	*
Atrial fibrillation	0.14	0.09	-0.05	-0.06	***
Congestive heart failure	0.24	0.14	-0.1	-0.11	***
Chronic kidney disease	0.32	0.24	-0.08	-0.11	***
Chronic obstructive pulmonary disease	0.31	0.25	-0.07	-0.11	***
Glaucoma	0.25	0.21	-0.04	-0.03	
Hip fracture	0.03	0.01	-0.02	-0.02	**
Hyperlipidemia	0.81	0.74	-0.07	-0.07	***
Hypertension	0.81	0.73	-0.08	-0.1	***
Benign prostatic hyperplasia	0.13	0.18	0.05	0	
Acquired hypothyroidism	0.31	0.21	-0.09	-0.07	***
Osteoporosis	0.24	0.14	-0.09	-0.06	***
Rheumatoid arthritis/osteoarthritis	0.58	0.51	-0.07	-0.05	*
Stroke/transient ischemia attack	0.13	0.06	-0.06	-0.07	***
Number of patients	157,074	852			

Notes: This table reports the differences in mean characteristics between trial participants and non-participants, in the Trial Participation Sample. The differences are adjusted for a patient's disease state (cancer type × stage at diagnosis × treatment line). Statistical significance is reported in the final column after Bonferroni correction. *** significant at the 1% level; ** significant at 5% level; * significant at the 10% level.

Table 3: LASSO Predictions of Serious Adverse Event Risk

A. LASSO Model Coefficients	
	Dependent variable:
Z-Standardized independent variables:	Average SAE Rate
Chronic kidney disease	0.0053
Kim frailty index	0.0050
Total chronic conditions	0.0035
Chronic obstructive pulmonary disease	0.0029
Hip fracture	0.0023
Hypertension	0.0018
Congestive heart failure	0.0018
Stroke/transient ischemia attack	0.0018
Anemia	0.0012
Acute myocardial infarction	0.0012
Benign prostatic hyperplasia	0.0006
Atrial fibrillation	0.0006
Rheumatoid arthritis/osteoarthritis	0.0004
Asthma	0.0004
Hyperlipidemia	0.0003
Ischemic heart disease	0.0000
Female	-0.0006
Osteoporosis	-0.0007
Glaucoma	-0.0008
Alzheimer's/dementia	-0.0009
Acquired hypothyroidism	-0.0009
Age^2	-0.0014
Sample size	23,437
B. LASSO Predictions of SAE Risk	
Mean	0.036
Standard deviation	0.019
10th percentile	0.013
25th percentile	0.022
50th percentile	0.033
75th percentile	0.047
90th percentile	0.062

Notes: Dependent variable is a patient's monthly average occurrence of any severe adverse event during months 0, 1, and 2 following treatment initiation. Predictor variables were Z-standardized and reported in order of the coefficient. Fixed effects for 20 disease states (defined by cancer type and stage at diagnosis)were included in the regression and prevented from shrinkage by LASSO regularization. The following coefficients were dropped after regularization: age, cataracts, depression, and diabetes. Age was centered at 65 years. LASSO is estimated in the SAE Sample, excluding patients who initiated treatment after February 2020 because they lack complete three-month follow-up.

Table 4: Effect of Treatment Initiation on Serious Adverse Events

	Dependent Variable: Serious Adverse Event			
	Model 1	Model 2	Model 3	Model 4
Post Treatment Initiation	0.024	0.024	0.024	0.024
	[0.021, 0.027]	[0.021, 0.027]	[0.021, 0.027]	[0.021, 0.027]
Post × (SAE Risk - $\overline{\text{SAE Risk}}$)		0.470		0.476
		[0.344, 0.549]		[0.349, 0.555]
Calendar time	X	X	X	X
Time since diagnosis	X	X	X	X
Individual fixed effects			X	X

Notes: This table reports results from estimation of Eq. 2, estimating the effect of treatment initiation on the monthly SAE rate for 0–2 months after treatment initiation. We apply an imputation estimator that controls for fixed effects in: month since diagnosis, calendar time, treatment group (based on timing of treatment initiation 3, 4, 5, or 6 months after diagnosis), and cancer type by stage. In Models 1-2, we also control for patient age and predicted SAE risk. In Models 3-4, we include individual fixed effects. Models 1 and 3 average the individual treatment effect estimates from equation 4, and Models 2 and 4 regress the individual treatment effect estimates on patient-level demeaned risk of SAE (using the jackknifed risk model), as in equation 5. The SAE Sample includes 24,002 unique patients and 448,017 patient × month observations from pre-treatment until up to 2 months post-treatment. Confidence intervals are calculated using a bootstrap procedure with 500 iterations, clustered at the patient level.

Table 5: Relationship between Trial Participation and SAE Risk

	Dependent Variable: Trial Participant		
	Full sample	Advanced stage cancers	Early stage cancers
Panel A. Controlling for Disease Cohorts Only			
SAE risk	-0.128	-0.248	-0.088
	(0.010)	(0.029)	(0.009)
Panel B. Controlling for Age, Sex, Race, and Disease cohorts			
SAE risk	-0.101	-0.233	-0.056
	(0.010)	(0.029)	(0.009)
Mean trial participation	0.005	0.011	0.004
Sample size	157,926	39,145	118,781

Notes: This table reports coefficients from linear regressions where the outcome variable is Trial Participation and the key independent variable of interest is predicted SAE risk, in the Trial Participation Sample. Panel A reports the coefficient estimate and standard error of the SAE risk variable in a linear regression controlling for 28 disease state fixed effects (cancer type × stage at diagnosis × treatment line × biomarkers). Panel B reports the results from a linear regression controlling also for age, age squared, sex, and race (White, Black, American Indian and Alaska Native, Asian and Pacific Islander, Other). Heteroskedastic robust standard errors are reported in parentheses.

ONLINE APPENDIX: Trials Avoid High Risk Patients and Underestimate Drug Harms

Jason Abaluck, Leila Agha, Sachin Shah

A1 Model Derivation

First, we define our procedure for removing γ-bias. Consider a new trial with choice probabilities s_j and define $\Delta_j = s_j - p_j$, the difference between these probabilities and the population probabilities for each subgroup. In principle, any vector of perturbations Δ_j to the subgroup shares satisfying $\sum_j \Delta_j = 0$ and $\sum_j \Delta_j \gamma_j = 0$ will produce a new design which is γ-unbiased. We will specifically choose the perturbation Δ^* which is the weighted-L^2-minimal adjustment to the existing trial probabilities r_j. That is, we find Δ_j satisfying the two above constraints and minimizing the expression $\sum_j \frac{(\Delta_j - (r_j - p_j))^2}{p_j}$. This yields:

$$\Delta_j^* = (r_j - p_j) - \frac{B_\gamma p_j}{Var_p(\gamma)}(\gamma_j - E_p(\gamma)) \tag{A1}$$

where $E_p(\gamma) = \sum_j p_j \gamma_j$ and $Var_p(\gamma) = \sum_j p_j (\gamma_j - E_p(\gamma))^2$. Note that this minimization problem is identical if we write: $\tau_j = \beta_k(\alpha_j + \delta\gamma_j)$, adding a trial-specific scaling factor so that trialists who know α_j and γ_j still do not know net impacts. To ease notation, we assume $\beta_k = 1$ below.

Let $h = \frac{Cov_p(\alpha,\gamma)}{Var_p(\gamma)}$, the population regression weighted coefficient of α on γ. Consider the case when $B_\alpha = \sum_j (r_j - p_j)\alpha_j$ (initial α-bias) and $B_\gamma = \sum_j (r_j - p_j)\gamma_j$ (initial γ-bias) have the same sign. The τ-bias of the original design is given by $B_\alpha + \delta B_\gamma$. The bias of the new design is given by:

$$\begin{aligned}
B_\tau(s) &= \sum_j (s_j - p_j)\alpha_j + \delta \cdot 0 \\
&= \sum_j [(r_j - p_j) + (s_j - r_j)]\alpha_j \\
&= B_\alpha - \frac{B_\gamma}{Var_p(\gamma)} \sum_j p_j(\gamma_j - E_p(\gamma))\alpha_j \\
&= B_\alpha - \frac{Cov_p(\alpha,\gamma)}{Var_p(\gamma)} B_\gamma \\
&= B_\alpha - hB_\gamma
\end{aligned} \tag{A2}$$

where $Cov_p(\alpha, \gamma) = \sum_j p_j(\gamma_j - E_p(\gamma))\alpha_j$. We want to know when:

$$|B_\alpha + \delta B_\gamma| \geq |B_\alpha - hB_\gamma| \tag{A3}$$

Let $x = \frac{B_\alpha}{B_\gamma}$. We can rewrite A3 as $|x - h| \leq |x + \delta|$.

Since $x + \delta > 0$ when B_α and B_γ have the same sign, this is equivalent to satisfying both of the following two inequalities concurrently:

$$x - h \leq x + \delta \rightarrow -h \leq \delta \rightarrow \delta \geq -h \tag{A4}$$

$$-(x - h) \leq x + \delta \rightarrow h - x \leq x + \delta \rightarrow \delta \geq h - 2x \tag{A5}$$

Finally, that is equivalent to:

$$\max(-h, h - 2\frac{B_\alpha}{B_\gamma}) \leq \delta \tag{A6}$$

Note that if $h = 0$ (meaning that α_j and γ_j are uncorrelated), this always holds.

A2 Cohort construction & variable definitions

Adverse event cohort construction. The study cohort included bladder, breast, colon, lung, pancreas, rectal, and renal cancer patients in the SEER-Medicare database who met the following inclusion criteria: first diagnosed with local, regional, or distant cancer between 2009 and 2020, aged between 66 and 100 years old at diagnosis, continuously enrolled in Medicare Parts A/B/D from 1 year before to 1 year after diagnosis, initiated treatment with any chemotherapy or immunotherapy between 3 months and 1 year after diagnosis. The sample includes patients with cancer diagnosis dates between January 2014 through December 2019. Linked Medicare claims extend back to January 2013, allowing us to record comorbidities and hospitalizations for 12 months prior to diagnosis. The sample is constructed to record monthly SAE hospitalizations for 12 months prior to diagnosis and 6 months afterwards. We end our sample period in February 2020 to exclude the Covid-19 pandemic.

Trial participation cohort construction. Following our previous study on trial participation

(Shah et al., 2025), a cohort of breast, lung, pancreatic, and renal cancer patients who met the following inclusion criteria was constructed from the SEER-Medicare database: first diagnosed with local, regional, or distant cancer between 2014 and 2019, aged 65 years or older at diagnosis, continuously enrolled in Medicare Parts A/B (and Part D if renal cancer patient) from 1 year before to 6 months after index date or until death. See Appendix Table A2 for details on included disease categories and inclusion criteria.

Cancer treatment. Carrier, Outpatient, and Part D event files were used to identify patients treated with a cancer drug. Cancer treatment was identified by HCPCS codes within the ranges J8501-J9999, C9000-9499, and Q0083-Q0085 and NDC codes for drugs categorized as a Chemotherapy or Immunotherapy in the SEER CanMED database.

Severe adverse events. We constructed a set of adverse events and ICD-9/10 codes that were used in previous studies of cancer treatment toxicity (Bishnoi et al., 2021; Bittoni et al., 2018; Du et al., 2005; Freedman et al., 2011; Gunturu et al., 2022; Hansen et al., 2014; Herbach et al., 2022; Hershman et al., 2013; Rashid et al., 2015; Reeder-Hayes et al., 2017; Sanoff et al., 2012; Wieder and Adam, 2023). A severe adverse event was defined as a hospitalization for which the primary diagnosis was a potential adverse event. See Appendix Table A3 for a summary of SAE types.

A3 Regulating SAE Risk and Net Impacts

We conduct a calibration exercise to assess whether the condition in equation 9 is likely to hold in practice, in the context of cancer drug trials. This requires us to estimate h, the relationship between heterogeneous harms and benefits of treatment. Our empirical work above estimated γ_p, the heterogeneous harms of cancer drugs. However, our quasi-experimental approach is not well-suited to estimating heterogeneous benefits of cancer drugs α_p, because survival benefits manifest over a longer time horizon and our approach leveraging variation in treatment initiation captures short-run differences in outcomes. Instead, we use our data to estimate cancer mortality risk, and assume that the drug's treatment benefits are proportional to this risk.

Calibration of heterogeneous treatment benefits proceeds in two steps. First, to estimate cancer mortality risk, we use the SEER data to predict mortality from cancer within 1 year of diagnosis.

Variables used for prediction include tumor size, the number of positive regional lymph nodes, and treatment indicators for surgical or radiation treatment, and we control for primary tumor site and stage at diagnosis (local, regional, or distant). As expected, patients with larger tumors and those with more positive lymph nodes have higher cancer mortality risk, while patients receiving surgical treatment have lower mortality risk. Second, we estimate the proportional reduction in mortality from cancer drugs using clinical trial data. Specifically, we match the top 46 most commonly prescribed drugs in our sample to the clinical trials described on their FDA labels. We identify 66 randomized clinical trials reporting a survival endpoint (not all drugs report trials with mortality endpoints, but some drugs report more than one trial).[12] Across these 66 trials, we calculate an average 20.3% reduction in mortality in the treatment group relative to the control group. In our SAE cohort, we estimate each patient's benefit of cancer drug treatment as a 20% reduction in their predicted 1-year cancer mortality risk.

To put our estimate of annual treatment benefits and monthly harms on a common cumulative scale, we scale up our monthly estimate of SAE treatment effects. Our baseline calibration multiplies the monthly SAE treatment effects by 3, to estimate the impact over the 3-month period following treatment initiation. This is conservative since many patients have longer durations of cancer drug treatment (beyond 3 months), and so continue to have elevated SAE hospitalization rates for more than 3 months; we compare this to alternative calibrations that multiply these SAE TEs by 6 or 12.

Next, using these measures of treatment benefits and harms, we estimate $Cov_p(\alpha, \gamma)$. We construct this measure using only variation in benefits and harms that arise conditional on the primary tumor site and cancer stage, since drug trials typically test for efficacy within a cancer site and stage. We find that $h = -0.0488$, suggesting only a weak relationship between SAE treatment effects and treatment benefits, whereby patients with a 10 percentage point (p.p.) greater probability of a drug-induced SAE event would experience a 0.5 p.p. greater decline in 1-year cancer mortality from treatment.

In our empirical work above, we presented evidence that B_γ is negative, i.e. trials understate the

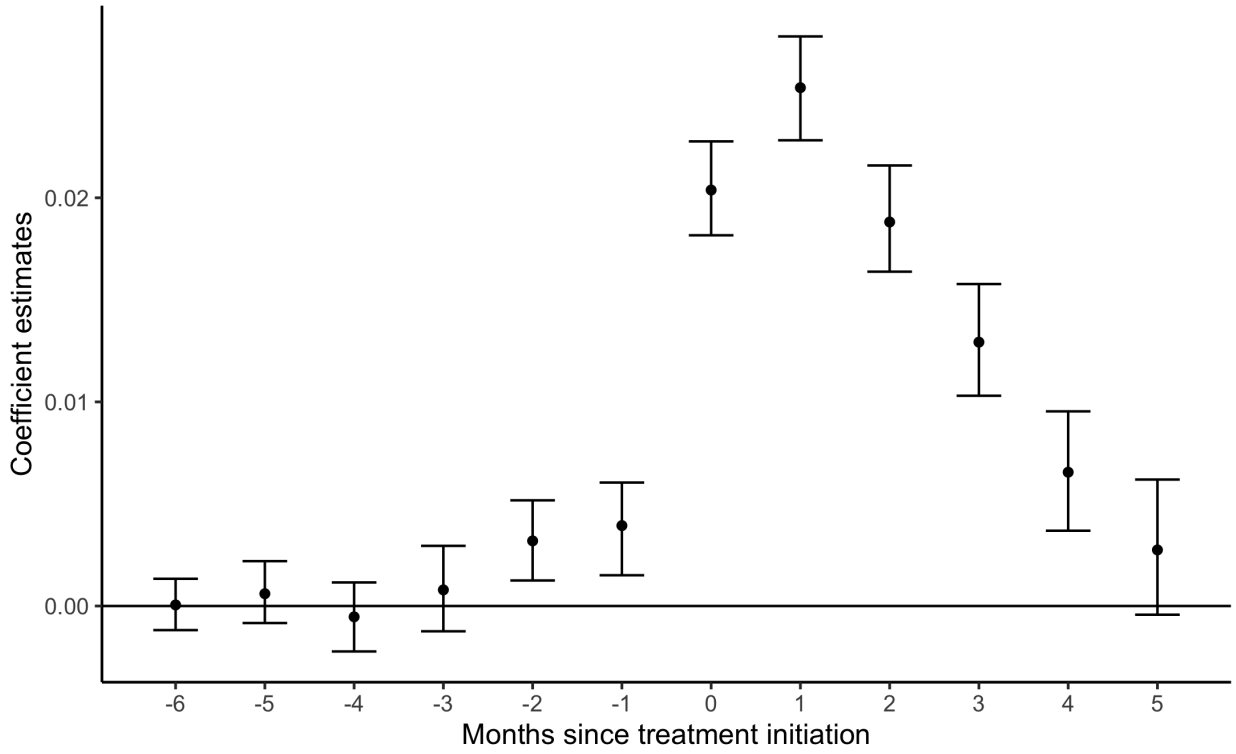
[12] Some trials do not directly report 1-year outcomes but show Kaplan-Meier survival curves that include a 1-year horizon; we use pixel-counting to estimate mortality at the 1-year horizon.

risk of drug-induced harms. We do not have direct evidence about the sign of B_α, but we suspect it is likely negative or zero if trials select patients who are likely to have greater benefits (consistent with trial sponsors' incentives to demonstrate that the drug is effective). Given that $h < 0$, if $B_\alpha \leq 0$, then comparing $-h = 0.049$ to δ is sufficient for checking the condition in equation 9. To calibrate δ, we turn to Shabaruddin et al. (2013), which summarizes evidence on the utility cost of drug-induced adverse events for cancer patients. The paper reports a mean utility with cancer of 0.88 (relative to 1 for healthy state and 0 if dead), and a utility decrement of -0.17 for the SAE hospitalization lasting 2–5 days, based on Beusterien et al. (2009). Thus, an SAE hospitalization corresponds to 0.193 times the utility cost of death ($0.193 = 0.17/0.88$). Since $-h = .049 < \delta = .193$, this implies that improving SAE representativeness would also improve representativeness in terms of net impacts. This remains true as long as SAE hospitalizations are at least 4.8% of the utility cost of death, but is no longer true below that threshold. If SAE hospitalizations are elevated for 6 rather than 3 months assumed at baseline, then $-h$ drops to $-h = .024$, thus suggesting improving SAE representativeness would improve net benefit representativeness unless SAE hospitalizations were less than 2.5% of the utility cost of death.

If B_α and B_γ have opposite signs, so that current trial enrollment patterns *understate* drug benefits, then the condition is met as long as δ is sufficiently large and the total bias in benefits is substantially smaller than the total bias in harms. If $\delta = .19$, then magnitude of bias from understating benefits B_α under current enrollment patterns would have to be 12% or less of the bias from understating harms B_γ. This proportion decreases as δ shrinks, as reported in Table A8.

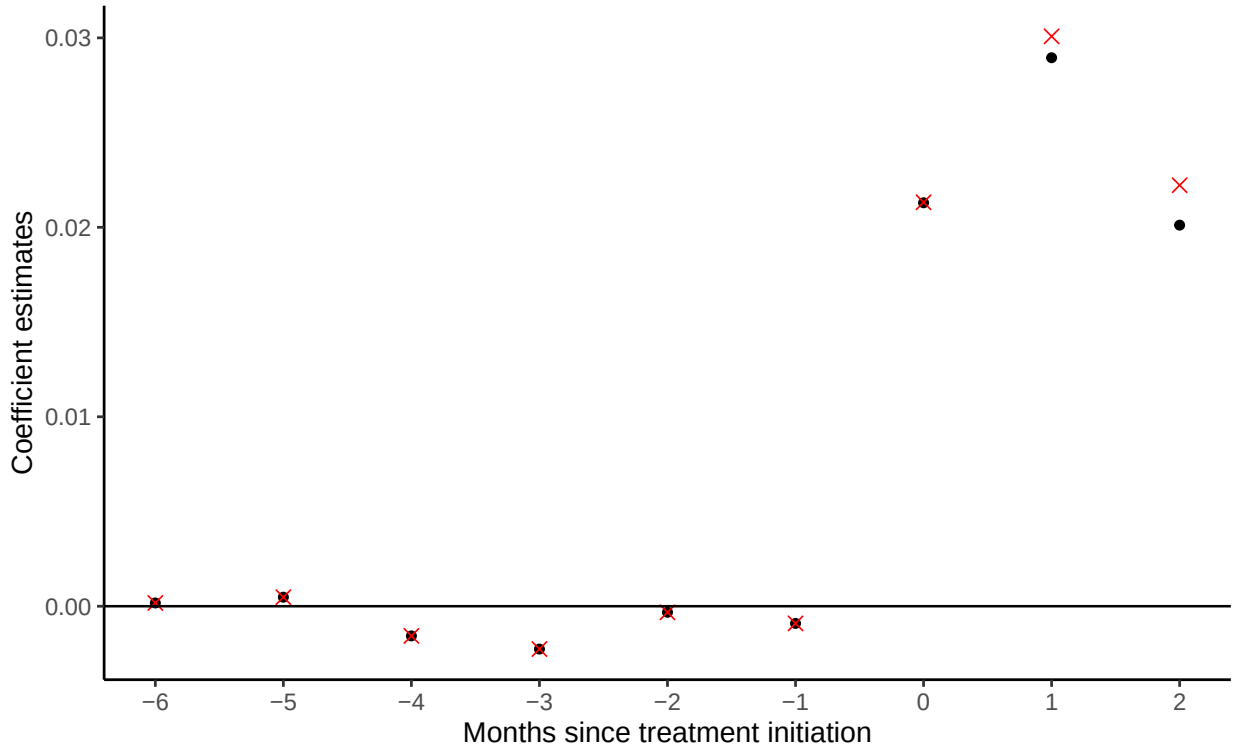
With sufficient data, the FDA could use analogous methods to the approach we develop here to try to regulate both α_j and γ_j representativeness concurrently. The analysis here suggests that, given reasonable assumptions about δ and B_α, regulating SAE representativeness alone is likely to improve net benefit representativeness in the setting of cancer drug trials.

Figure A1: Serious Adverse Event Rate, Before & After Treatment Initiation: Broad Cohort with Treatment Initiation in Months 3-11



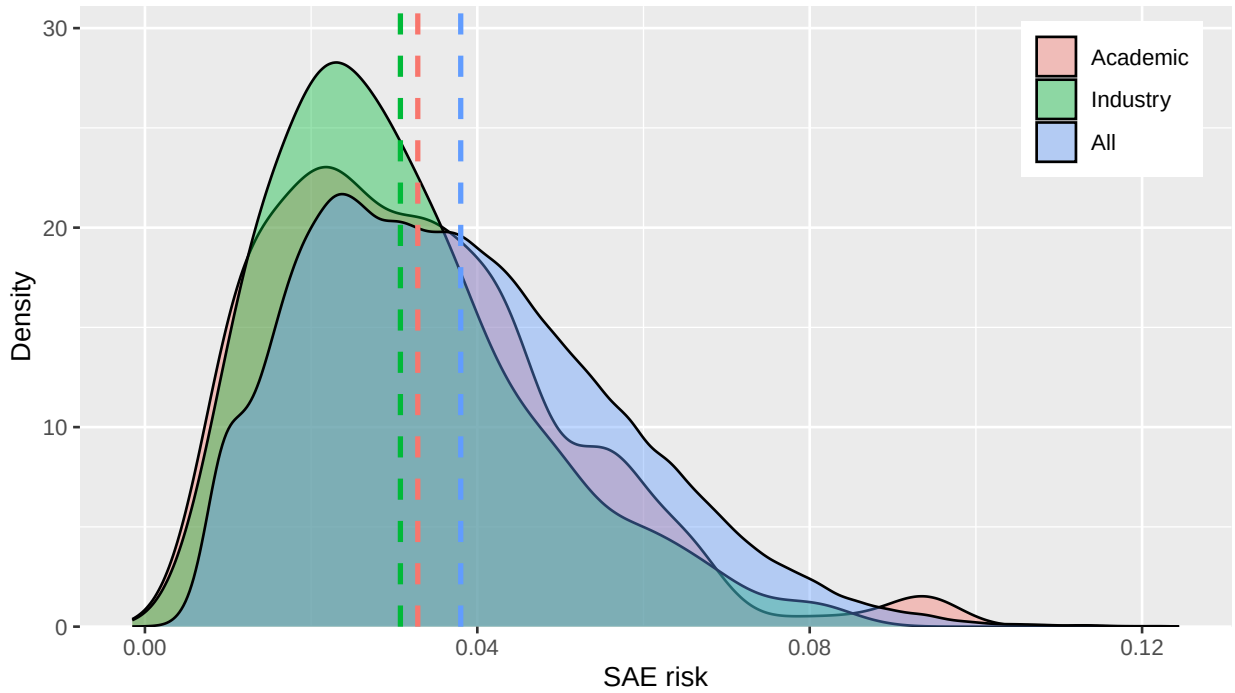
Notes: This figure reports the coefficents $\hat{\theta}^R$ from Equation 6 using a broader cohort of patients than the baseline analysis. Specifically, this analysis expands to include all patients who initiate treatment within 3-11 months following diagnosis (instead of restricting to cohorts initiating treatment 3-6 months following diagnosis, as in the main analysis).

Figure A2: Serious Adverse Event Rate, Before & After Treatment Initiation: Censoring Patients after Death



Notes: This figure reports the coefficients $\hat{\theta}^R$ from Equation 6 in the main SAE sample described in the text, initiating treatment 3-6 months after diagnosis. The black dots reflect our baseline estimates, while the red X's display results from an alternative sample that drops patients from the sample in months after their death. 3.47% of patients who initiated treatment in months 3-5 died within 5 months since diagnosis.

Figure A3: Distribution of SAE Risk Among Trial Participants vs. Full Cohort by Trial Sponsor



Notes: This figure displays kernel density plots of the SAE risk distribution among industry- and academic-sponsored trial participants vs. the full trial-eligible cohort. We categorize trials in our sample according to the sponsor listed on ClinicalTrials.gov. Patients are re-weighted to match the distribution of specific disease cohorts for the pooled sample of trial participants (including all sponsors); for disease cohort definitions see Appendix Table A2. Mean risk equals 0.0307 for industry-sponsored trial participants, 0.0328 for academic-sponsored trial participants, and 0.0380 in the full cohort. The Trial Participation sample includes 157,926 patients.

Table A1: SEER Registries

California
Connecticut
Georgia
Hawaii
Idaho
Iowa
Kentucky
Louisiana
Massachusetts
Detroit (Metropolitan)
New Jersey
New Mexico
New York
Texas
Utah
Seattle (Puget Sound)

Notes: This table lists the geographic sites of all the SEER Registries included in our analysis.

Table A2: Trial-eligible cohort disease state definitions (Part 1)

Cancer Type	Neoadjuvant/Adjuvant	Advanced
Breast	**Trials for neoadjuvant/adjuvant breast cancer were included.** **Patients**: We included all patients with local or regional disease at diagnosis. We constructed 8 cohorts based on 4 biomarker possibilities (HER2+, HR+, HER2+/HR+, triple negative) and 2 stage possibilities (local, regional). Index date is diagnosis date.	**Trials for advanced-stage breast cancer were not included** because significant heterogeneity in therapeutic options and criteria for clinical trial eligibility for patients with advanced-stage disease increases the risk that we misattribute patients as potential trial participants who do not have the disease state studied in the trial. **Patients**: Excluded patients with distant disease at diagnosis.
Lung	**Trials for neoadjuvant/adjuvant lung cancer were not included** because many trials were specifically for patients who did or might get surgical resection, while others were for patients getting chemoradiation, and registry data do not provide imaging data necessary to identify patients appropriate for surgery vs radiation. Thus, there would be higher misattribution than in other neoadjuvant/adjuvant categories. **Patients**: Excluded patients with local or regional disease at diagnosis and no drug claims qualifying them for the advanced-stage cohort (as defined in the "Advanced" column).	**Trials for advanced-stage lung cancer were included.** **Patients**: We constructed 4 separate cohorts for Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC), differentiating first vs. second line treatment cohorts. Patients must have received a drug from the relevant drug list. Drug list for NSCLC*: pemetrexed, paclitaxel, paclitaxel protein-bound, etoposide, docetaxel, gemcitabine, pembrolizumab, nivolumab, vinorelbine, atezolizumab Drug list for SCLC*: etoposide, irinotecan, atezolizumab, lurbinectedin, topotecan First line: If diagnosed with distant disease, start any listed drug on or after diagnosis date. If diagnosed with local or regional disease, then first-line treatment requires drug claim to begin $\geq$180 days after diagnosis. Second line: Start a new drug after the final claim of first-line drug(s). Index date is the earlier of the date of treatment initiation (for relevant treatment line) or first report of relevant RCT participation. * Note: platinum chemotherapies are excluded from these drug lists as they are generally used in conjunction with other drugs and are sometimes re-used in subsequent lines of therapy, increasing the risk of misattribution of line of therapy.

Table A2: Trial-eligible cohort disease state definitions (Part 2)

Cancer Type	Neoadjuvant/Adjuvant	Advanced
Pancreatic	**Trials for neoadjuvant/adjuvant pancreatic cancer were included.** **Patients**: Any patients with local or regional disease at diagnosis. Index date is diagnosis date.	**Trials for advanced-stage pancreatic cancer were included.** **Patients**: We differentiated first vs. second line treatments using claims data as described below. Patient must be receiving a drug from the relevant drug list (see below). Patients must have distant disease at diagnosis. Drug list: paclitaxel, paclitaxel protein-bound, gemcitabine, irinotecan, oxaliplatin, capecitabine, fluorouracil. First line: Start any listed drug on or after diagnosis date. Second line: Start a new drug after the final claim of first-line drug(s). Index date is the earlier of the date of treatment initiation (for relevant treatment line) or first report of relevant RCT participation.
Renal	**Trials for neoadjuvant/adjuvant renal cancer were included.** **Patients**: Any patients with local or regional disease at diagnosis. Index date is diagnosis date.	**Trials for advanced-stage renal cancer were included.** **Patients**: We studied only first-line treatments. Patients must be receiving a drug from the drug list (see below). Patients must have regional or distant disease at diagnosis. Drug list: sunitinib, cabozantinib, nivolumab, ipilimumab, pazopanib, axitinib, everolimus, lenvatinib, sorafenib, temsirolimus. First line: Start any listed drug on or after diagnosis date. Index date is the earlier of the date of treatment initiation or first report of relevant RCT participation.

Table A3: Proportion of patients with SAE within 3 months of treatment initiation, by SAE Type

Any	0.0371
Infection	0.0224
Kidney	0.0064
Cardiac	0.0036
Hematologic	0.0034
Gastrointestinal	0.0030
Liver	$\leq$0.0004
Lung	$\leq$0.0004

Notes: This table lists the rate of having at least one hospitalization for a potential SAE of each listed subtype. The sample includes 24,002 unique patients and 69,140 patient-month observations from 0–2 months from treatment initiation.

Table A4: Effect of Treatment Initiation on SAE Rate:
Broad Cohort with Treatment Initiation in Months 3-11

	Dependent Variable: Serious Adverse Event			
	Model 1	Model 2	Model 3	Model 4
Post-period average	0.022	0.022	0.022	0.022
	[0.020, 0.023]	[0.020, 0.023]	[0.020, 0.024]	[0.020, 0.024]
Post-period risk effect		0.447		0.463
		[0.350, 0.497]		[0.363, 0.509]
Calendar time	X	X	X	X
Time since diagnosis	X	X	X	X
Individual fixed effects			X	X

Notes: This table reports the coefficients $\hat{\theta}^R$ from Equation 6 using a broader cohort of patients than the baseline analysis. The broadened sample includes 30,589 unique patients and 597,916 patient-month observations spanning from 12 months before diagnosis until up to 2 months post-treatment initiation.

Table A5: Relationship between Trial Participation and SAE Risk: Logistic Regression

	Dependent Variable: Trial Participant		
	Full sample	Advanced stage cancers	Early stage cancers
Panel A. Controlling for Disease Cohorts Only			
SAE risk	-28.349	-26.875	-30.047
	(2.320)	(3.265)	(3.302)
Panel B. Controlling for Age, Sex, Race, and Disease cohorts			
SAE risk	-22.923	-25.111	-19.626
	(2.371)	(3.315)	(3.338)
Mean trial participation	0.005	0.011	0.004
Sample size	157,926	39,145	118,781

Notes: This table reports coefficients from logistic regressions where the outcome variable is Trial Participation and the key independent variable of interest is predicted SAE risk. Panel A reports the coefficient estimate and standard error of the SAE risk variable in a linear regression controlling for 28 disease state fixed effects (cancer type × stage at diagnosis × treatment line × biomarkers). Panel B reports the results from a logistic regression controlling also for age, age squared, sex, and race (White, Black, American Indian and Alaska Native, Asian and Pacific Islander, Other). Heteroskedastic robust standard errors are reported in parentheses.

Table A6: Relationship between SAE Risk and Trial Participation

Percentile	Predicted SAE risk	Predicted SAE treatment effect	Predicted trial participation
10th	0.0144	0.0136	0.0081
25th	0.0213	0.0169	0.0073
50th	0.0328	0.0223	0.0058
75th	0.0474	0.0292	0.0039
90th	0.0618	0.0359	0.0021

Notes: This table reports the distribution of Predicted SAE risk, Predicted SAE treatment effects, and trial participation in the Trial Participation Cohort. Predicted SAE treatment effects represent monthly estimates obtained using Model 2 coefficients from Table 4. Predicted trial participation obtained using the coefficient estimates from the linear probability model estimated in Table 5 Panel A Column 1.

Table A7: Predicted SAE Risk by Time-to-Treatment Initiation

Months since diagnosis until treatment	Sample size	Mean predicted SAE risk
0	11,171	0.0395
1	30,241	0.0407
2	23,363	0.0377
3	12,509	0.0363
4	5,924	0.0358
5	3,444	0.0346
6	2,125	0.0340

Notes: The risk model summarized in Table 3 was applied to a subset of patients who initiated treatment between 0 and 6 months since diagnosis.

Table A8: Calibration of Model: Does improving SAE representativeness improve net benefit representativeness?

	Monthly SAE TE ×3	**Monthly SAE TE ×6**
	Values of h:	
	-0.0488	-0.0244
Values of δ	*Max value of $\lvert B_\alpha \rvert / \lvert B_\gamma \rvert$, if $B_\alpha > 0$:*	
0.01	0.029	0.017
0.05	0.049	0.037
0.10	0.074	0.062
0.20	0.124	0.112
0.30	0.174	0.162
0.40	0.224	0.212

Notes: This table summarizes the calibration exercise reported in Section 6. We consider two alternative re-scalings of our monthly SAE TE results to match the annual cancer mortality risk variable. Column 1 assumes only three months of elevated SAE risk (since that was the post-period used in estimation) and multiplies our monthly estimate by 3. Column 2 assumes six months of elevated SAE risk. The top row reports the slope coefficient from a regression of treatment benefits (predicted mortality reduction) on treatment harms (predicted rate of induced SAE hospitalizations), estimated in the SAE sample. If $B_\alpha <= 0$ so that under current enrollment patterns, trials overstate treatment benefits, then checking that $-h < \delta$ is sufficient to ensure that the proposed regulation would improve SAE representativeness. If $B_\alpha >= 0$ (trials understate treatment benefits), then we also require the condition reported in the bottom panel: i.e. that B_α is much smaller in magnitude than B_γ, with the maximum proportion depending on δ and reported in the bottom table.