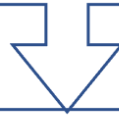


A. Initial effect estimates: 9,460 drug combinations tested for effect on up to 26 cancers, for a total 179,937 estimates (**Supplementary Data 1**)

Step	Bias accounted for
Drug-wide	• Time under exposure • High-dimensional set of confounders



B. Evaluations: Test consistency with true effect for 9460 drug combinations and 668 combinations of drug A + ingredient B (shared by >1drug) (**Supplementary Data 2, 3**)

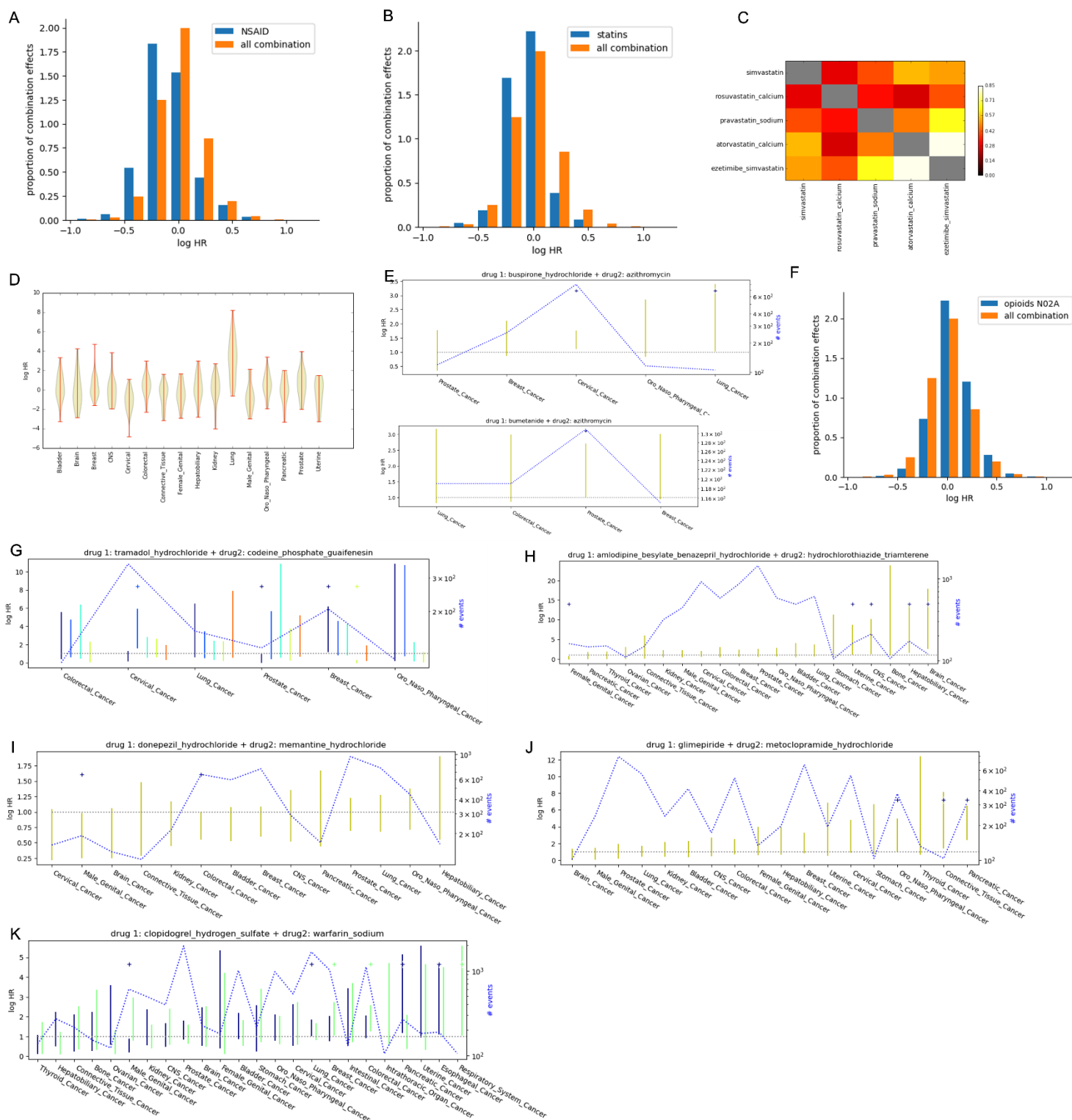
Step	Bias accounted for
Consistency of effects (hierarchical model)	• Confounding specific to single cancer • Confounding specific to a single drug (signal across drugs sharing an ingredient)
Specificity (negative controls)	• Combinations associated with general unwellness or susceptibility to disease



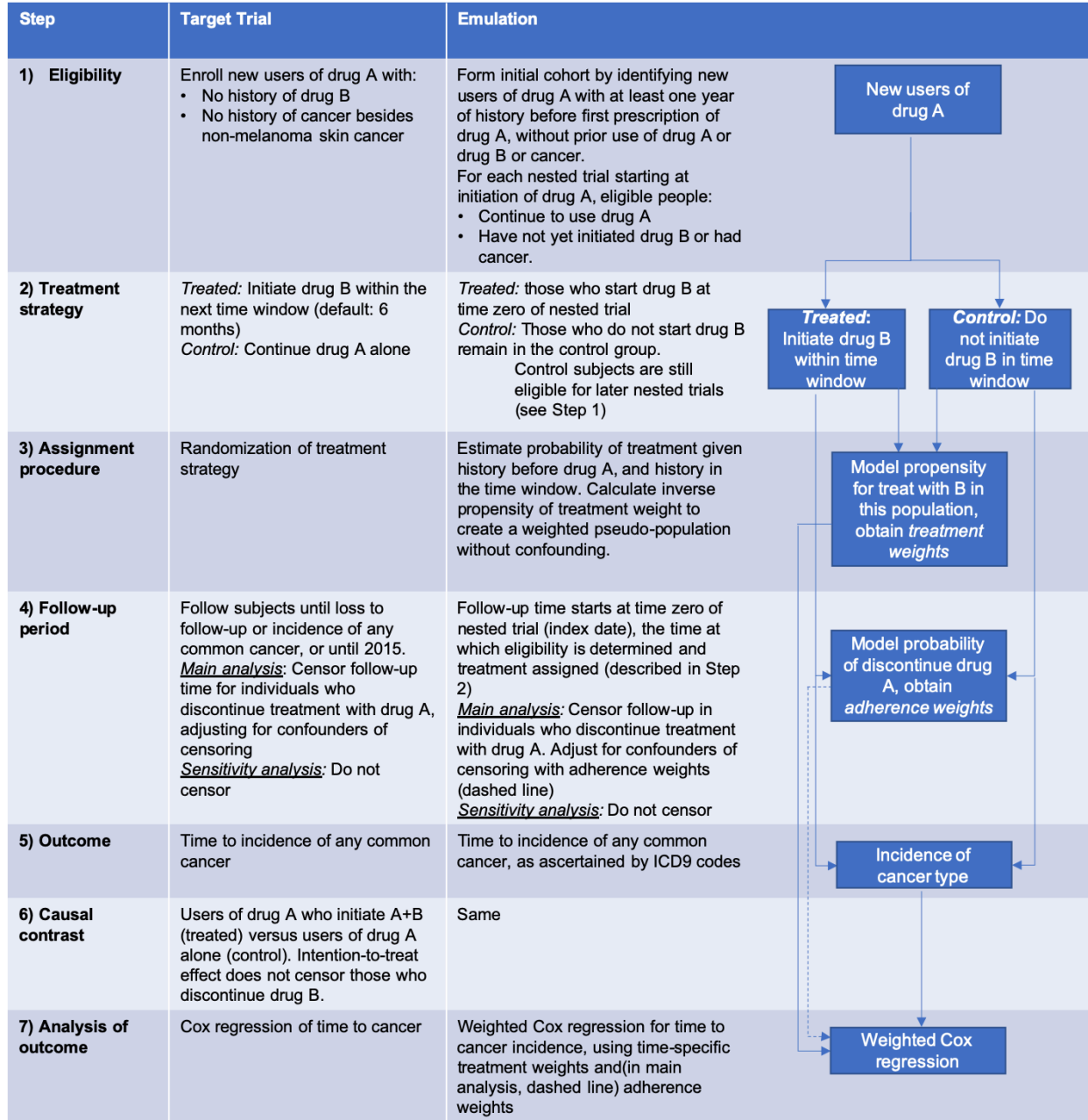
C. Sensitivity analysis: Performed on combinations passing all tests above (**Supplementary Tables 2, 3**)

Parameter	Sensitivity analysis
Reverse causality tests	• Missing data on cancer diagnosis accounting for association
Width of time windows	• Vary discretization of time into windows (3 month, 6 month)
Definition of exposure to combination	• With and without censoring people who discontinue first drug
Amount of drug	• Partition second drug by amount: does higher amount have stronger effect?

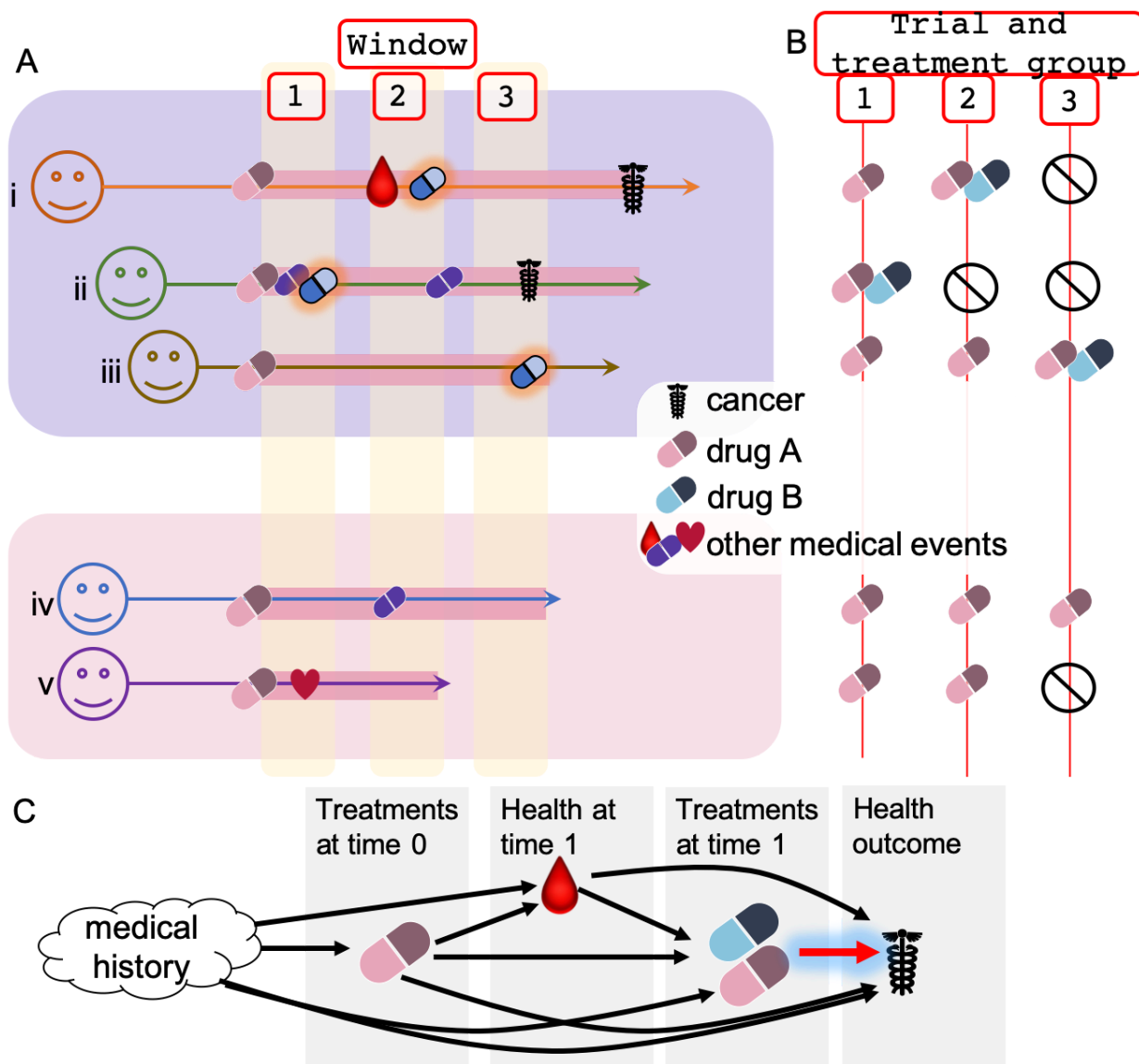
Supplementary Fig S1 Summary of analyses. **A.** We perform a drug combination-wide screen for effects. **B.** Then, we perform evaluations of the per-combination effects. **C.** Finally, we perform sensitivity analyses for combinations passing the evaluations.



Supplementary Fig S2 Comparison to reported drug signals. **A**. Distribution of estimated drug effects per combination including NSAIDs, compared to all combinations (ranksum test p-value = 5e-21). **B**. Distribution of estimated drug effects per combination including statins, compared to all combinations (ranksum test p-value = 1e-9). **C**. Correlation of statin drug effects, compared to each other statin involving a combination combined with the same drug. **D**. Distribution of estimated drug effects per combination from combinations involving a COPD drug (ATC category R03BB), showing elevated association with lung cancer. **E**. Two combinations involving azithromycin (ATC J01FA), replicating Stør's finding of increased risk (log HR posterior 98% CI = [.01, .6], and [.05, .45] respectively, Supplementary Data 3). **F**. Distribution of estimated drug effects per combination including opioids, compared to all combinations, showing elevated risk of cancers (ranksum test p-value = 8e-27). **G**. Combination with opioid expectorant (R05FA) replicating Stør's finding of increased risk (each colored line represents one replicate cohort study) (log HR posterior 98% CI = [.007-.54], Supplementary Data 3). **H**. Combination involving hydrochlorothiazide with increased risk of cancer (log HR posterior 98% CI = [.03-.66], Supplementary Data 3). **I**. Combination of anticholinesterase (dementia drugs) (log HR posterior 98% CI = [-0.36 -0.06], Supplementary Data 2). **J**. Propulsive (A03FA) combination (log HR posterior 98% CI = [.06-.7], Supplementary Data 3). **K**. Anticoagulant combination (log HR posterior 98% CI = [.08-.5], Supplementary Data 3).



Supplementary Fig S3 Summary of emulation of a target trial for testing drug combination effect on cancer incidence.



Supplementary Figure S4. Enrollment for trial and addressing confounding. **A.** Medical history for five people exposed to drug A (pink pill) is represented with time along the horizontal axis. Follow-up time is divided into 3 windows. The top three people are eventually exposed to the combination of drugs A and B (blue pill), while the bottom two people remain in the comparator group (drug A alone). **B.** Representation of the expansion of person-time for these five people into a sequence of randomized trials. Person 1 is unexposed at time window 1 (trial 1), and thus is in the comparator group, but still eligible to participate in trial 2. At trial 2 they receive the combination, and thus are assigned to the joint-exposure group in trial 2, and they are not eligible to participate in trial 3. **C.** Each arrow represents influence or possible influence of one event in health history on another event. We assume that all items in health history can influence what happens later in time. The main edge of interest is the one highlighted in red, and all other edges represent nuisance associations. To remove confounding edges, we model probability of treatment at any time point t : $p(\text{exposed}_t | \text{health history before } t)$

Supplementary Table 1: Expansion into trials

Person	Trial (time zero)	Follow- up index (calendar time = time zero + follow-up - 1)	Treatment	Outcome	Treatment weight model	Treatment probability	comment
i	1	1	A	no	yes	$p(A+B_{i,1}=False)$	person i is not eligible for trials after trial 2 because they took the combination at trial 2
i	1	2	A	no	no	$p(A+B_{i,1}=False)$	
i	1	3	A	no	no	$p(A+B_{i,1}=False)$	
i	1	4	A	yes	no	$p(A+B_{i,1}=False)$	
i	2	1	A+B	no	yes	$p(A+B_{i,1}=False) \times p(A+B_{i,2}=True)$	
i	2	2	A+B	no	no	$p(A+B_{i,1}=False) \times p(A+B_{i,2}=True)$	
i	2	3	A+B	yes	no	$p(A+B_{i,1}=False) \times p(A+B_{i,2}=True)$	
v	1	1	A	no	yes	$p(A+B_{v,1}=False)$	person v is only followed for 2 windows, they have no drug B or outcome in either.
v	1	2	A	no	no	$p(A+B_{v,1}=False)$	
v	2	1	A	no	yes	$p(A+B_{v,1}=False) \times p(A+B_{v,2}=False)$	

Supplementary Model:

```
data{
  ///
  int <lower=1> N; // tot
  int <lower=1> Noutcome;
  real cancer_effect_sd; // prior variance for mean of eff
  real percancer_devsd; // prior variance for mean of eff
  real across_cancers_sd; // prior variance for mean of eff
  vector[N] expvals;
  vector[N] se;
  //int<lower=0, upper=1> outcome_is_cancer[Noutcome];
  int N_class;
  int<lower=1,upper=N_class> drug2[N];

  //int<lower=0, upper=1> cancer_indicator[N];
  int<lower=1, upper=Noutcome> outcome_id[N];

}
parameters{
  vector[N_class] eff_per_drug_cancer_tilde[Noutcome];
  real class_cancer_effect;
  real<lower=0> class_allcancer_sd;

  vector<lower=0>[Noutcome] outcome_sd_across_drugs; // for each outcome, the spread across
drugs might be different
  vector[Noutcome] class_per_cancer_effect_tilde;
}
transformed parameters{
  real Xeff_rep[N];
  vector[N_class] eff_percancer[Noutcome];
```

```

    vector[Noutcome] class_per_cancer_effect = class_per_cancer_effect_tilde * class_allcancer_sd
+ class_cancer_effect;
    for(i in 1:Noutcome){
        eff_percancer[i] = eff_per_drug_cancer_tilde[i] * outcome_sd_across_drugs[i] +
class_per_cancer_effect[i];

    }
    for (i in 1:N){
        Xeff_rep[i] = eff_percancer[outcome_id[i]][drug2[i]] ;
    }
}

model {
    // these 2 are the effect & var of effect overall of cancer of drugs in that group
    class_cancer_effect ~ normal(0, cancer_effect_sd);
    class_allcancer_sd ~ normal(0,across_cancers_sd); // overall, how wide is the spread across
cancers

    // then these are the variations across cancers of that
    class_per_cancer_effect_tilde ~ normal(0, 1);

    // then we get to the per-drug variation from that for each cancer: the random deviations,
and what we scale these by per cancer
    outcome_sd_across_drugs ~ normal(0, percancer_devsd);
    for(i in 1:Noutcome){
        eff_per_drug_cancer_tilde[i] ~ normal(0,1); // each of these being random spread around
the per-cancer effect, across drugs in that class
    }

    if(run_estimation==1){
        expvals ~ normal(Xeff_rep, se);    }
}

```

