How Often: All-by-All

- OHDSI Global Symposium: proof of concept for large-scale incidence rate generation
- All by all: incidence rates for **every drug, every side effect**
  - About 40,000,000 pairs
  - Plus every side effect (disease prevalence)
  - Stratified

Why All-by-All Incidence?

- Planning – how often does a disease happen
- Side effects – if a drug causes a side effect, how often does it happen
  - Many side effects are well known, but most clinicians have no idea of the incidence
- Detection – compare adverse event reports of new treatment with baseline rates

Why All-by-All Incidence?

- Incidence rate calculation is “simple”
  - Feasible to execute all-by-all
  - Fewer assumptions
  - (But still more complicated than it looks!)
- Characterization = non-causal, but can still be useful
  - If incidence is low and side effect is not serious, then we’re good
  - If incidence is high, then need to look out for it even if not caused by drug

Why All-by-All: Fill Gaps in Evidence

- What currently exists?
  - Prospective studies, RCTs
    - Potentially causal
    - Hard to do at scale (they're not doing 40 million estimates)
    - Small sample size (not rare)
    - Less generalizable
  - Drug package inserts
    - Not causal (may have nothing to do with the drug)
    - Few have rates
  - Guessing

Why All-by-All: Fill Gaps in Evidence

- What currently exists?
  - Prospective studies, RCTs
    - Potentially causal
    - Hard to do at scale (they're not doing 40 million estimates)
    - Small sample size (not rare)
    - Less generalizable
  - Drug package inserts
    - Not causal (may have nothing to do with the drug)
    - Few have rates
  - Guessing

All-By-All incidence rates attempt to fill in these gaps

Why All-by-All: Assess reliability of outcomes

- Estimate EHRs and claims measurement error
  - Differences in outcomes between databases
    - One of several factors
  - Differences between RCT or prospective and EHR or claims
    - RCT >> EHR could be lack of coding
  - Averaged across drugs

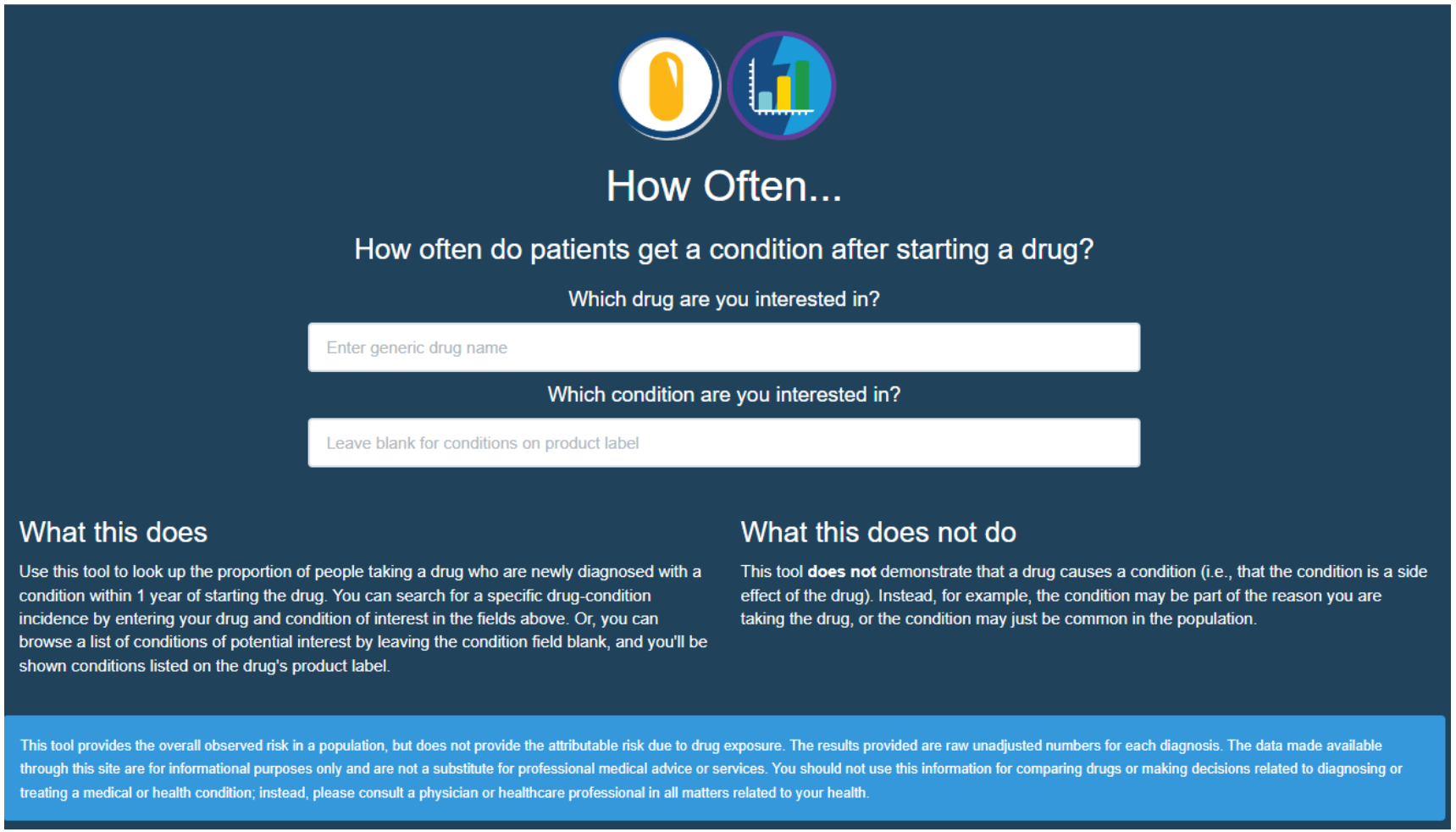
Caveats to incidence rates

- Why rates might be high:
  - Causal (attributable risk), *or*
  - Indication is a risk (ex: statins reduce risk for MI, but the indication for statins is that they are at high risk for MI)
- 40 million estimates even before stratification
  - 2600 drugs, 18000 SNOMED outcomes
  - Neet to ensure rates are not being misinterpreted (i.e., causal) when used

Uncertainty > Accuracy

- The task is not so much better incidence rates but better estimation of uncertainty
  - Don’t need to get each of 40 million estimates right
  - Need to get proper uncertainty estimates for them
- And then conveying that uncertainty to readers
- Work on improving the rates in parallel

All-by-All Demo Went Live in 2017



Experience with the All-by-All

- Columbia-NYP emergency department, QA
  - Used by some staff on patients with unexplained symptoms
  - No formal evaluation
- Was the trigger for Anna Ostropolets's Data Consult Service
  - Angioedema with penicillin
- Elise Ruan found that it was reliable for several examples but some exceptions
  - sexual dysfunction and sertraline; muscle pain and atorvastatin
- Did not maintain proof of concept

We need you!

- We have shown proof-of-concept
- Time to produce the all-by-all with stratification
  - Need to determine what strata we can do
- Phenotypes
  - Condition codes, phenotype library
  - New methods in between – research
- How to aggregate the information
- How to communicate (non-causal) incidence rates
- We need your help
  - Will be iterative and will want to keep it up to date