Hypothesized versus Observed Effect Sizes in Chemoprevention Trials – A Systematic Review

Jens Eickhoff[1], Jen Birstler[1], Guanhua Chen[1], Zhumin Zhang[1], Vikrant Sahasrabuddhe[2], Eva Szabo[2], and KyungMann Kim[1]
[1]University of Wisconsin-Madison, Madison, WI, [2]National Cancer Institute, Bethesda, MD

Department of Biostatistics and Medical Informatics
UNIVERSITY OF WISCONSIN
SCHOOL OF MEDICINE AND PUBLIC HEALTH

CP-CTNetDMACC
Cancer Prevention Clinical Trials Network
Data Management, Auditing, and Coordinating Center

Background and Rationale

- Early phase efficacy chemoprevention trials are typically designed to demonstrate feasibility, and signals of efficacy
- Many studies, however, fail to detect hypothesized effect sizes
- The *primary objective* of this systematic review was to conduct a in-depth evaluation for failure to detect chemoprevention intervention effects
- The following study elements were evaluated:
  - Hypothesized vs. observed effect sizes
  - Planned vs. actual sample sizes
  - Post-hoc power analyses

Materials and Methods

- Chemoprevention trials under the Chemoprevention Consortia Program of the Division of Cancer Prevention, National Cancer Institute between 2003 and 2019 were reviewed
- *Inclusion Criteria*: Single or multi-arm efficacy/biomarker trials
- *Exclusion Criteria*:
  - Dose finding or safety studies
  - Pharmacokinetic/pharmacodynamics studies
  - Bioequivalence studies
  - Non-inferiority/equivalence studies

Data Extraction

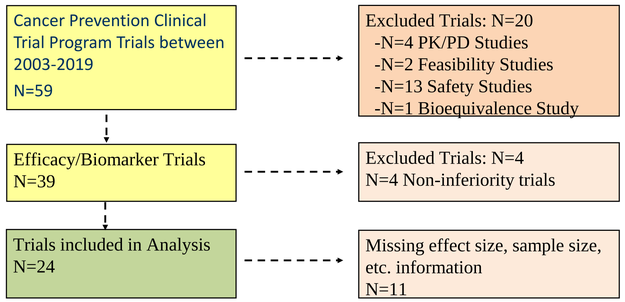
- **Study protocols** were reviewed to gather information regarding:
  - Study design (randomized vs. non-randomized)
  - Analysis populations (intent-to-treat vs. per-protocol)
  - Statistical analysis plan
  - Statistical Hypothesis
  - Sample size calculation (effect sizes, power)

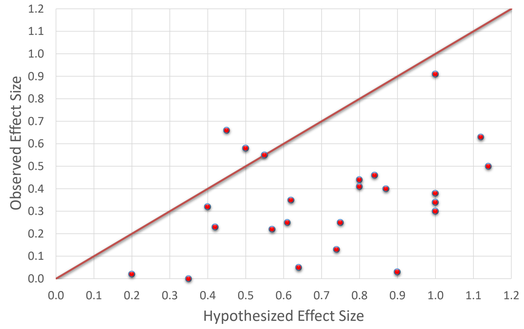
- **Manuscripts** and **study reports** were reviewed to gather information regarding:
  - Observed sample size
  - Observed effect size
  - Accrual period
  - Statistical analysis
  - Early termination
  - Recruitment duration

Analysis

- Observed effect sizes were standardized, i.e., ES=$|\mu_{Trt}-\mu_{Ctrl}|/SD$
- In cases where only p-values were reported, the observed effect sizes were estimated using a normal approximation method based on the p-values and sample sizes
- Differences between the hypothesized vs. observed effect sizes were calculated
- Concordance between hypothesized vs. observed effect sizes was evaluate using the intra-class correlation (ICC) coefficient
- Post-hoc power calculation based on the observed effect sizes were conducted
- Subgroup analyses were conducted (single arm vs. multi-arm studies, intent-to-treat vs. per-protocol, accrual goal achieved vs. not achieved)

Results

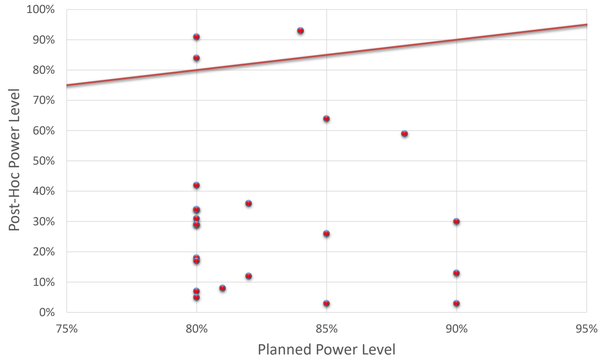




- Hypothesized ES: Median: 0.75 vs. Observed ES: Median 0.34
- Median Difference: 0.37 (IQR 0.19-0.60)
- Hypothesized ES were *smaller* than observed ES in 92% of the studies
- Low Concordance between hypothesized vs. observed ES: ICC: 0.18

Results

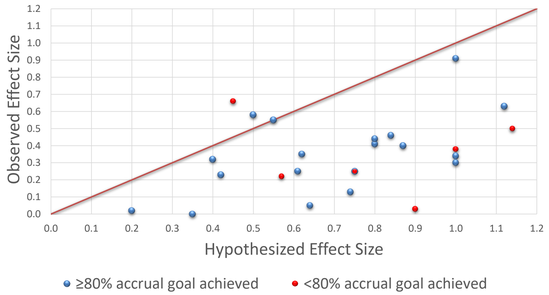
Post-Hoc Power Analysis



- Median post-hoc power level was only 29%
- The post-hoc power exceeded the planned power in only 2 trials

Subgroup Analysis: Accrual Goal Achieved vs. Not Achieved

	Hypothesized ES	Observed ES	Difference
≥80% accrual goal achieved	0.70 ± 0.25	0.35 ± 0.23	0.35 ± 0.23
<80% accrual goal achieved	0.81 ± 0.29	0.36 ± 0.24	0.45 ± 0.41



Subgroup Analysis: Intent-to-treat (ITT) vs. Per-Protocol Analysis (PP)

	Planned ES	Observed ES	Difference
ITT	0.72 ± 0.26	0.35 ± 0.23	0.37 ± 0.28
PP	0.73 ± 0.27	0.36 ± 0.21	0.36 ± 0.25

Subgroup Analysis: Randomized trials vs. single arm trials

	Hypothesized ES	Observed ES	Difference
Randomized	0.78 ± 0.29	0.32 ± 0.25	0.46 ± 0.22
Single arm	0.56 ± 0.08	0.41 ± 0.23	0.15 ± 0.30*

*p<0.05 for randomized vs. single arm difference

- For single arm trials, the mean hypothesized ES was close to the observed ES

Conclusions

- For the majority of early phase cancer prevention efficacy trials, the observed effect sizes were substantially smaller than the hypothesized effect sizes
- Sample size calculations should be conducted under realistic assumptions regarding anticipated effect sizes
- Sample size calculations need to balance potential detectable/clinical important effect sizes that can be realistically accrued with the need to detect effect sizes to justify subsequent large scale confirmatory trials

Funding

NIH 1U24 CA242637-01: *CP-CTNet Coordinating Center*
NIH/NCI P30CA014520