CADTH 25 YEARS

ANNIVERSARY LECTURE SERIES

25 years of supporting informed health care decisions.

 cadth.ca | @cadth_acmtes
Hollow Hunt for Harms

Jacob Stegenga

University of Utah

jacob.stegenga@utah.edu
Harms of medical interventions are systematically underestimated.

Two fundamental problems with clinical research:

1. Trial design
2. Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

**First in Human, Never Seen Again**

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

**Clinical Trials and the Abuse of Power**

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Phases of Clinical Research
Overview

**Operationalizing Harm**

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Some Orwellian terminology

Medical interventions have many effects
  •  labels list on average 70 side effects
  •  some drugs have over 500 known side effects
  •  much data on harms is of ‘adverse events’

‘drug safety’
  •  *signal* of a *safety finding*, in a *safety report*
Operationalizing harms contributes to their underestimation
Operationalizing harms contributes to their underestimation

Pediatric suicidal ideation caused by some antidepressants (ADs)
• Early meta-analyses: ADs do not cause suicidal ideation in children

➢ Data from Hamilton Rating Scale for Depression (HAMD)
  HAMD question on suicidality:
  0 = Absent
  1 = Feels life is not worth living
  2 = Wishes he were dead or any thoughts of possible death to self
  3 = Suicidal ideas or gesture
  4 = Attempts at suicide

• HAMD insensitive to significant changes in suicidality
A curious aside

HAMD question on ‘insight’:
0 = Acknowledges being depressed and ill
1 = Acknowledges illness but attributes causes to bad food, climate, overwork, virus, need for rest, etc
2 = Denies being ill at all
Rosiglitazone: running example

Rosiglitazone, made by GSK, was the world’s leading drug for type 2 diabetes

By 2005, evidence that rosiglitazone causes cardiovascular disease and death

GSK funded a trial to disprove this

• Primary outcome: *composite* of hospitalizations and deaths
• outcome rate between groups made to appear similar
• hospitalization is a social event & trial took place in many countries
• added variability mitigated apparent difference between groups
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Phases of Clinical Research

- **Preclinical**: Plausibility of treatment (20-80 people)
- **Phase-I**: Safety (100+ people)
- **Phase-II**: Efficacy & Safety (1000+ people)
- **Phase-III**: Efficacy & Dosing
- **Phase IV**: Long-term effectiveness

Laboratory testing | Human trials
Publication bias of phase 1 trials

First test of a novel drug in humans

Risky for subjects
  • e.g. CD28-SuperMAB

Foundation for assessing harms

Vast majority unpublished

$\rightarrow \textit{systematic underestimation of harms}$
Rosiglitazone: publication bias

Rosiglitazone: modulates peroxisome proliferator-activated receptors (PPARs)

Over 50 PPAR modulators have failed clinical tests, many because of harms

“few data on toxicity are available in the public domain because of the common industry practice of not publishing safety findings for failed products.” (Nissen and Wolski 2007)
Rosiglitazone: knowledge of mechanisms

Knowledge of how drug works ought to influence estimation of harm profiles

PPAR modulators regulate the expression of many dozens of genes

“effects of these agents are unpredictable and can result in unusual toxicities”

Evidence-based medicine (EBM) downplays mechanistic reasoning
Publication bias of phase 1 trials

Opacity of phase 1 evidence is systematically skewed

Ok... so, we should expect drugs that wrongly appear safe after phase 1 trials, but then come to be seen as harmful, as more evidence is gathered

Just among PPAR modulators, many have been withdrawn*
  - troglitazone - causes liver damage
  - tesaglitazar - causes elevated serum creatinine
  - pioglitazone - causes bladder cancer
  - muraglitazar - causes heart attacks, strokes, and death

*in some jurisdictions
Overview

Operationalizing Harm

First in Human, Never Seen Again

**Clinical Trials and the Abuse of Power**

Jump Now, Look Later

Secrecy
Phases of Clinical Research

- **Preclinical**: Plausibility of treatment
  - 20-80 people

- **Phase-I**: Safety
  - 100+ people

- **Phase-II**: Efficacy & Safety
  - 1000+ people

- **Phase-III**: Efficacy & Dosing

- **Phase IV**: Long-term effectiveness
Trials designed to be sensitive to benefits, not harms

Statistical power:
- the probability of avoiding a ‘type II’ error (wrongly concluding no difference between experimental group and control group)

Power, more broadly:
- the sensitivity of a trial to detect an effect of an intervention
Distinguishing $\text{power}_B$ and $\text{power}_H$

Power usually refers to ability of a trial to detect benefit: $\text{power}_B$

We also ought to be concerned with ability of a trial to detect harms: $\text{power}_H$

$\text{power}_B$ and $\text{power}_H$ trade-off against each other

Researchers almost always try to maximize $\text{power}_B$ at expense of $\text{power}_H$

- Financial incentive to avoid type II$_B$ error and commit type II$_H$ error
Subject selection maximizes $\text{power}_B$ at expense of $\text{power}_H$

Many trials include only subjects who:
- are most likely to benefit
- are similar to one another

Many trials exclude subjects who:
- are least likely to benefit
- have comorbidities
- are on other drugs
- fare poorly on test drug
- are non-compliant on test drug
- respond to placebo
Trials designed to be sensitive to benefits, not harms

“enrichment strategies”

“run-in periods”

“randomized withdrawal design”

“sequential parallel comparison design”

Trial designs advance to overcome bitter pill of placebo effect

Monica Heger
Trials designed to be sensitive to benefits, not harms

‘run-ins’ maximize $\text{power}_B$ at the expense of $\text{power}_H$

E.g. 15 trials analyzed by FDA regarding AD use in children
  • only 3 showed benefits
  • 2 of these 3 were of fluoxetine
  • problem: these trials employed placebo run-ins
  • all children put on placebo for 1 week & responders excluded
Effect of power$_B$-maximizing strategies: *subjects* are different from *patients*

Such differences are known to influence harm profile of drugs.

Older people, pregnant women, and patients on other drugs are more likely to be harmed by drugs, but they are also excluded from trials.

E.g. most common harm of statins is myopathy (myalgia to rhabdomyolysis)

- risk is higher among women, elderly, and people with comorbidities
- precisely the kinds of people excluded from trials

Trials designed to be sensitive to benefits, not harms
Trials designed to be sensitive to benefits, not harms

Two factors that contribute to trade-off: trial size and duration

Many harms are rarer than benefit, and occur later than benefit

Trials enroll enough subjects to achieve good power \( B \), and no more
- any more subjects would increase cost
- any more subjects would increase chance of detecting harms
- trial size is set to optimize power \( B \) with no regard for power \( H \)

Trials extend for a duration long enough to detect benefit, and no longer
- any longer would increase cost
- any longer would increase chance of detecting harms
- e.g., AD studies often last only a few weeks
- trial duration optimizes power \( B \) against power \( H \)
Trials designed to be sensitive to benefits, not harms

Subject withdrawal from trial may contribute to underestimation of harms

Insufficient reporting of subject withdrawals is ubiquitous

E.g. review of 133 publications of RCTs published in 2006 in six top journals
  • no information on severe adverse events in 27%
  • no information on subject withdrawal due to harms in 47%
Rosiglitazone: power$_B$ versus power$_H$

A meta-analysis suggested that rosiglitazone causes an increased risk of heart attack and death (Nissen & Wolski 2007)

The individual trials were small and did not have sufficient power$_H$ to show this effect

GSK funded RECORD trial to refute this alleged causal relation

- 7 inclusion criteria
- 16 exclusion criteria
- 99% of the subjects were Caucasian
- subjects in trial were healthier than broader population
  - e.g. heart attack rate of 4.5 per 1000 person years, about 40% of the equivalent target population
Publication bias of trials distorts research record

E.g.: Meta-analysis with access to all data on reboxetine

- data on 74% of patients was unpublished
- 7 trials compared reboxetine to placebo
  - 1 had positive results: published
  - 6 had null results (10 times as many patients): all unpublished
Example: paroxetine

SSRI made by GSK, prescribed ‘off-label’ for children

From 1994 – 2002, GSK conducted numerous trials of paroxetine in children

None showed benefit

Not published

“It would be commercially unacceptable to include a statement that efficacy had not been demonstrated, as this would undermine the profile of paroxetine.”
<table>
<thead>
<tr>
<th>Type of Outcome</th>
<th>Number of Outcomes</th>
<th>Outcomes with Complete Information, n (Percent(^a))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Not Publicly Available</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CSR(^b) (n = 101)</td>
</tr>
<tr>
<td>All outcomes(^d)</td>
<td>1,080</td>
<td>930 (86)</td>
</tr>
<tr>
<td>Benefit outcomes</td>
<td>456</td>
<td>385 (84)</td>
</tr>
<tr>
<td>Mortality</td>
<td>92</td>
<td>92 (100)</td>
</tr>
<tr>
<td>Clinical events</td>
<td>119</td>
<td>108 (91)</td>
</tr>
<tr>
<td>Symptoms</td>
<td>215</td>
<td>168 (78)</td>
</tr>
<tr>
<td>HRQoL</td>
<td>30</td>
<td>17 (57)</td>
</tr>
<tr>
<td>Harm outcomes</td>
<td>624</td>
<td>545 (87)</td>
</tr>
<tr>
<td>AEs</td>
<td>101</td>
<td>93 (92)</td>
</tr>
<tr>
<td>SAEs</td>
<td>101</td>
<td>89 (88)</td>
</tr>
<tr>
<td>Withdrawal due to AEs</td>
<td>101</td>
<td>92 (91)</td>
</tr>
<tr>
<td>Special AEs(^a)</td>
<td>321</td>
<td>271 (84)</td>
</tr>
</tbody>
</table>
Publication bias is ubiquitous

<table>
<thead>
<tr>
<th>Drug</th>
<th>Drug</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sertraline</td>
<td>Valdecoxib</td>
<td>Zidovudine</td>
</tr>
<tr>
<td>Citalopram</td>
<td>Gabapentin</td>
<td>Zalcitabine</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>Aprotinin</td>
<td>Acyclovir</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>Anti-arrhythmics</td>
<td></td>
</tr>
<tr>
<td>Reboxetine</td>
<td>Nesiritide</td>
<td></td>
</tr>
<tr>
<td>Mirtazapine</td>
<td>Rosiglitazone</td>
<td></td>
</tr>
<tr>
<td>Duloxetine</td>
<td>Ezetimibe</td>
<td></td>
</tr>
<tr>
<td>Lamotrigine</td>
<td>Simvastatin</td>
<td></td>
</tr>
<tr>
<td>Quetiapine</td>
<td>Cerivastatin</td>
<td></td>
</tr>
<tr>
<td>Rofecoxib</td>
<td>Oseltamivir</td>
<td></td>
</tr>
</tbody>
</table>
Phases of Clinical Research

- **Laboratory testing**
  - **Preclinical**
    - Plausibility of treatment
  - **Phase-I**
    - Safety
    - 20-80 people
  - **Phase-II**
    - Efficacy & Safety
    - 100+ people
  - **Phase-III**
    - Efficacy & Dosing
    - 1000+ people
  - **Phase IV**
    - Long-term effectiveness

- **Human trials**
FDA only requires two RCTs that show some benefit, regardless of:

- how many RCTs have been performed
- publication bias
- effectiveness of competitor drugs
- low power $H$
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

**Jump Now, Look Later**

Secrecy
Phases of Clinical Research

- **Laboratory testing**
- **Human trials**
  - Preclinical
    - Plausibility of treatment
    - Safety
  - Phase-I
    - 20-80 people
    - Safety
  - Phase-II
    - 100+ people
    - Efficacy & Safety
  - Phase-III
    - 1000+ people
    - Efficacy & Dosing
  - Phase-IV
    - Long-term effectiveness
Post-approval data underestimates harms

After approval, harms are hunted with passive surveillance
  • scant prior evidence on harms, so patients are unwitting guinea pigs

Passive surveillance severely underestimates harms
  • physicians do not report commonly known harms
  • physicians do not attribute causality to drug
  • one empirical estimate: 94% of harms not reported
  • post-market studies confounded by health of compliant drug users
Post-approval data underestimates harms

Post-approval data on harms rarely come from randomized studies

• EBM gurus: “if a study wasn’t randomised, we suggest that you stop reading it and go on to the next article in your search”
• The FDA “is dominated by a world-view that believes only RCTs provide useful and actionable information and that postmarketing safety is an afterthought.”

So, EBM gurus & FDA denigrate vast majority of evidence on harms
Overview

Operationalizing Harm

First in Human, Never Seen Again

Clinical Trials and the Abuse of Power

Jump Now, Look Later

Secrecy
Evidence of harms is shrouded in secrecy

Publication bias systematically distorts the research record

Clinical researchers are bound by gagging clauses in contracts

Policy-makers often powerless or complicit
Example: ranibizumab

Christopher McCabe (CADTH talk, Sept 25, 2014):
• ranibizumab (Lucentis) is a ‘system-level lemon’

Document on ranibizumab submitted to NICE (U.K. National Institute for Clinical Excellence):
Adverse events: Adverse events were common but most were mild to moderate. Serious ocular events were rare in the MARINA and ANCHOR trials. Incidences of severe ocular inflammation varied between treatment arms, and were highest in the 0.5mg ranibizumab groups. The rate of serious ocular adverse events was

in the ranibizumab plus PDT group compared with

PDT. Endophthalmitis was reported by very few patients in the active treatment arms of the ranibizumab trials and none in the control arms. The condition occurred in up to 1.4% of 0.5mg dose ranibizumab patients in the ANCHOR trial, and the rate per injection was 0.05% in the MARINA trial. Endophthalmitis occurred in

of patients across the

and

trials.

Very few deaths were reported in the ranibizumab trials, with numbers of deaths being
Rosiglitazone: secrecy of evidence on harms

Several trials suggested that rosiglitazone causes cardiovascular harms

Steve Nissen requested all relevant data from GSK
  • GSK refused
  • But, GSK had been forced to develop a registry of trial data
  • Nissen got data from 42 RCTs of rosiglitazone
  → analysis: rosiglitazone increases cardiovascular events by 43%

• reviewer at *NEJM* faxed a copy of the paper to GSK
• internal email by director of research at GSK:
  “FDA, Nissen, and GSK all come to comparable conclusions regarding increased risk for ischemic events, ranging from 30% to 43%!”

*GSK and FDA already knew!*
Evidence of harms is shrouded in secrecy

Another example:

Researchers reviewing diet drugs orlistat and rimonabant tried to get unpublished data on harms from European Medicines Agency (EMA)

EMA refused, invoking commercial interests and patient confidentiality

EU ombudsperson rejected EMA arguments for secrecy

Researchers finally were sent documents on rimonabant...
60 page document on harms of rimonabant provided by EMA
Why are regulators complicit in such secrecy?

“regulatory capture”?

“FDA is inherently biased in favor of the pharmaceutical industry. It views industry as its client, whose interests it must represent and advance. It views its primary mission as approving as many drugs it can, regardless of whether the drugs are safe or needed.”
Rosiglitazone: defending secrecy of harm data

John Buse (1999) suggested that rosiglitazone may have cardiovascular risks

GSK initiated campaign to silence Buse (“Avandia Renegade”)
- threatened lawsuits in talks with Buse and Buse’s department chair
- GSK head of research:
  “sue him for knowingly defaming our product even after we have set him straight as to the facts ... launch a well planned offensive...”

Buse’s response: “please call off the dogs.”

By 2007, FDA estimated rosiglitazone caused 83000 heart attacks
Acknowledgements
- Banting Postdoctoral Fellowships Program (SSHRC)
- Institute for the History and Philosophy of Science and Technology, University of Toronto
- Ben Goldacre
- Heather Douglas, Alex Broadbent, Jonathan Fuller

Disclosures
- None

Caveats
- The material presented here contains no specific medical advice.
Harms of medical interventions are systematically underestimated.

Thank you for your attention.

Jacob Stegenga

jacob.stegenga@utah.edu
CADTH 25 YEARS

ANNIVERSARY LECTURE SERIES

25 years of supporting informed health care decisions.

cadth.ca | @cadth_acmats