



Data and Analytics for Human Health at NASA

Robert J. Reynolds, MS MPH PhD

Mortality Research & Consulting, Inc.

Translational Research Institute for Space Health, Baylor College of Medicine

National Aeronautics and Space Administration

August 18, 2020



About me



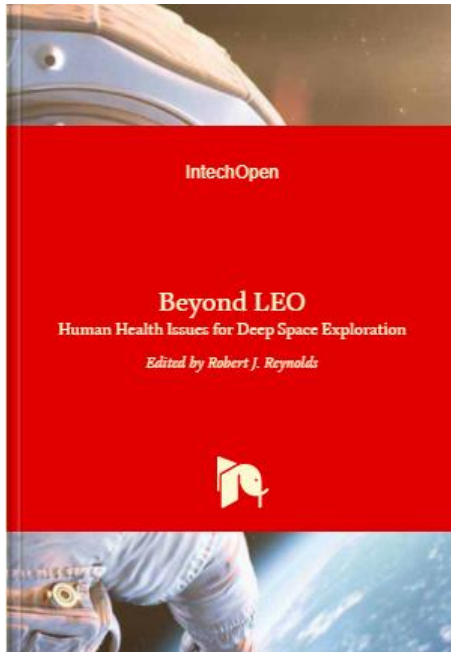
- MPH and PhD Epidemiology; MS Statistics; training in Data Science (particularly machine learning)
- 20+ years at universities, health insurance industry, statistical consulting, independent research
- Visiting Data Scientist at Human Health and Performance Directorate at NASA Johnson Space Center
- I have no financial conflicts of interest to report, and am not promoting a product or service in this talk
- All opinions expressed here are my own and are not representative of NASA or any other person or entity



Outline



1. Introduction to human health and performance data at NASA
2. Analytic challenges inherent to the NASA HHP data
3. Proposed solutions to the challenges
4. Examples of solutions in practice
5. Summary and conclusions



Introductory Chapter: Research Methods for the Next 60 Years of Space Exploration

By Robert J. Reynolds and Mark Shelhamer

<https://www.intechopen.com/books/beyond-leo-human-health-issues-for-deep-space-exploration/introductory-chapter-research-methods-for-the-next-60-years-of-space-exploration>



Human Health and Performance at NASA



- Dedicated to making spaceflight safer for human beings through mitigation of the hazards of spaceflight
- Human Health and Performance Directorate (HHPD) is composed of labs, clinical and operational medicine, research program, occupational health program, and more
- Researches and communicates the risks of human spaceflight to other parts of the Agency to guide requirements and standards





Astronaut health data at HHPD



Astronaut cohort as of 2020 (from Reynolds database):

- 59 years of human spaceflight
- 350 unique US astronauts
- 9,600+ US astronaut person-years
- 70% have <1 month in space

Other astronaut data:

- Hundreds of small-scale studies performed over last 50
- Routine medical care data from NASA clinic at JSC
- External medical data for retirees

Exposure data:

- Spacecraft exposure data from Shuttle and ISS - not protected, but still not easy to obtain; not cleaned for epidemiological analyses



NASA astronaut Scott Kelly wearing a headset and a Chibis Lower Body Negative Pressure (LBNP) Suit, undergoes ultrasound measurements for the Fluid Shifts experiment. (NASA Image: ISS045E015466)



Problem 1: Low statistical power



- Astronaut cohort small, with only 350 US astronauts ever selected
- Small exposure density of space travel makes health hazards uncertain
- Difficult to get a statistically significant result - good or bad - even when we expect one

$$Var(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

$$t = \frac{Z}{\sigma_{hat}/\sqrt{n}}$$



Problem 2: Violations of assumptions



- Statistical models require samples to be “iid” - independent identically distributed
- In an experimental setting this is achieved through randomization
- Astronauts are not randomly selected to the Astronaut Corps, to particular missions, or to particular duties
- Astronauts can choose to participate in research or not
- Violation of iid can make test statistics invalid and can lead to **confounding by indication**



Image by PierreSelim; used under CC BY 3.0 license
(<https://creativecommons.org/licenses/by/3.0/>)



Problem 3: Interpretation



- If we manage to obtain a statistically significant effect, how should we interpret it given the potential for confounding by indication?
- If we fail to see any significant effects where we believe we ought to *a priori*, how do we interpret that? Lack of effect or sample size issues?
- Does our assessment of the meaning of such results change with larger or smaller variance in our sample? Edge cases?



Problem 4: Preference for error



“The [NHST] framework itself is centered on the concept of errors in inference: when we can, we design our studies to avoid Type I error while simultaneously trying to limit Type II errors. In so doing we may make these errors—rather than what we might learn from our data—the primary consideration of our scientific activity. It should come as no surprise that when we make avoidance of error our top priority, we fail to learn all we can from our data.”

(Reynolds & Shelhamer 2020)



Solution 1: Realigning epistemology



- Think deeply about what it means for something to be *causal* (instead of merely “predictive” or “associative”)

Bradford Hill's Guidelines for Causation

1. Plausibility (face validity)
2. Coherence (laboratory & epidemiology)
3. Temporality* (cause precedes effect)
4. Strength (of association, effect size)
5. Biological gradient (dose-response)
6. Consistency (replication)
7. Specificity (of circumstances)
8. Experiment (clinical trial)
9. Analog (similar to known effects)

Bonus: Reversibility

* = necessary for causation

Major assumptions of modern causal inference

- **Positivity** – all units must have a non-zero chance of receiving all exposures
- **Consistency** – effects of exposure are not subject-dependent and are the same for all
- **Exchangeability** – exposure and covariates are independent

Analytic techniques in causal inference

- Propensity score analysis
- Marginal structural models
- G-computing

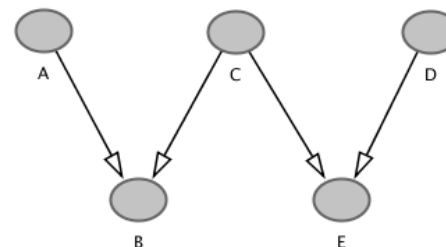


Solution 1: Realigning epistemology



➤ Directed acyclic graphs (DAGs)

- Network maps that reflect causal relationships
- DAG structure implies what is (and is not) possible in the system being modeled - before any data even collected



Other useful things to do with a DAG:

Community finding - clustering to find related risks
Ego networks - tracing causal families

Factorize joint prob. dist. over the DAG → Bayesian Network
Factorize corr. matrix over the DAG → Structural equation models

- $A \perp C$
- $A \perp D$
- $A \perp E$
- $B \perp D$
- $B \perp E \mid C$
- $C \perp D$



A is independent of C, i.e., they should have a correlation in-distinguishable from zero if this DAG is the correct causal model

Make your own DAGs:

<http://dagitty.net/>



Solution 1: Realigning epistemology



➤ Alternative hypotheses

- Equivalence testing - effect is inside a prespecified equivalence margin
- Non-Inferiority testing - effect is above a prespecified non-inferiority margin
- *Modus tollens* - “denying the consequent”, contrapositive reasoning

Excellent visualization for equivalence and non-inferiority testing:

<https://rpsychologist.com/d3/equivalence/>

Modus tollens general form

1. If P, then Q
2. Not Q
3. Therefore, not P

https://en.wikipedia.org/wiki/Modus_tollens



Solution 2: Alternative analytics



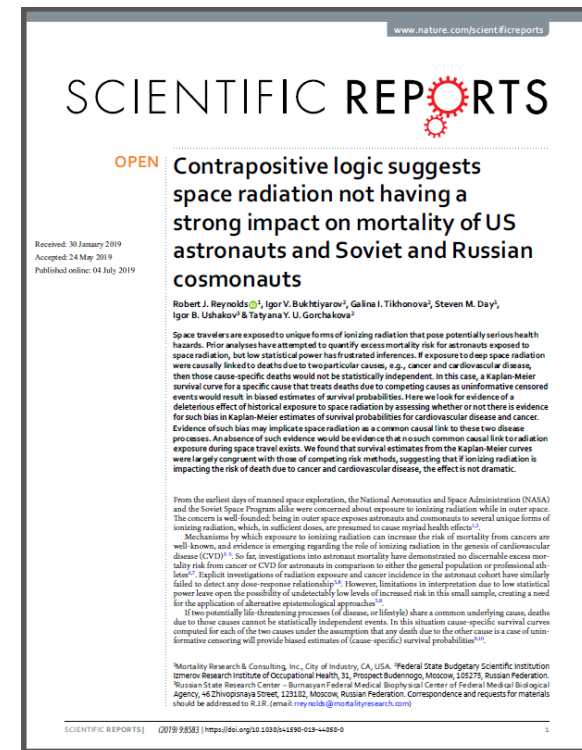
- **Resampling methods:** bootstrap
- **Simulation:** MCMC, agent-based
- **Bayesian methods** - weave in analog-based priors
- **Data Science** - Unsupervised ML and advanced data visualization
- **Alternative uses of common statistical models:** use properties of statistics other than formal tests to draw conclusions



Example 1: Space radiation and mortality



- Have historical doses of space radiation increased mortality rates for US astronauts?
- If so, this should be detectable in either CVD, cancer, or some other natural cause of death
- Low power may mean that these differences would be impossible to declare as meaningful under traditional analytic methods
- Prior investigations using conventional approaches already demonstrated no increases in risk from these causes of death





Example 1: Space radiation and mortality



Modus Tollens for mortality risk from space radiation:

If P then Q	IF doses of space radiation absorbed by astronauts are sufficient to increase risk of death for both CVD and cancer THEN survival curves from death by CVD and cancer will not be statistically independent.
Not Q	Survival curves from death by CVD and cancer ARE statistically independent THEREFORE
Not P	doses of space radiation absorbed by astronauts are sufficient to increase risk of death for both CVD and cancer.

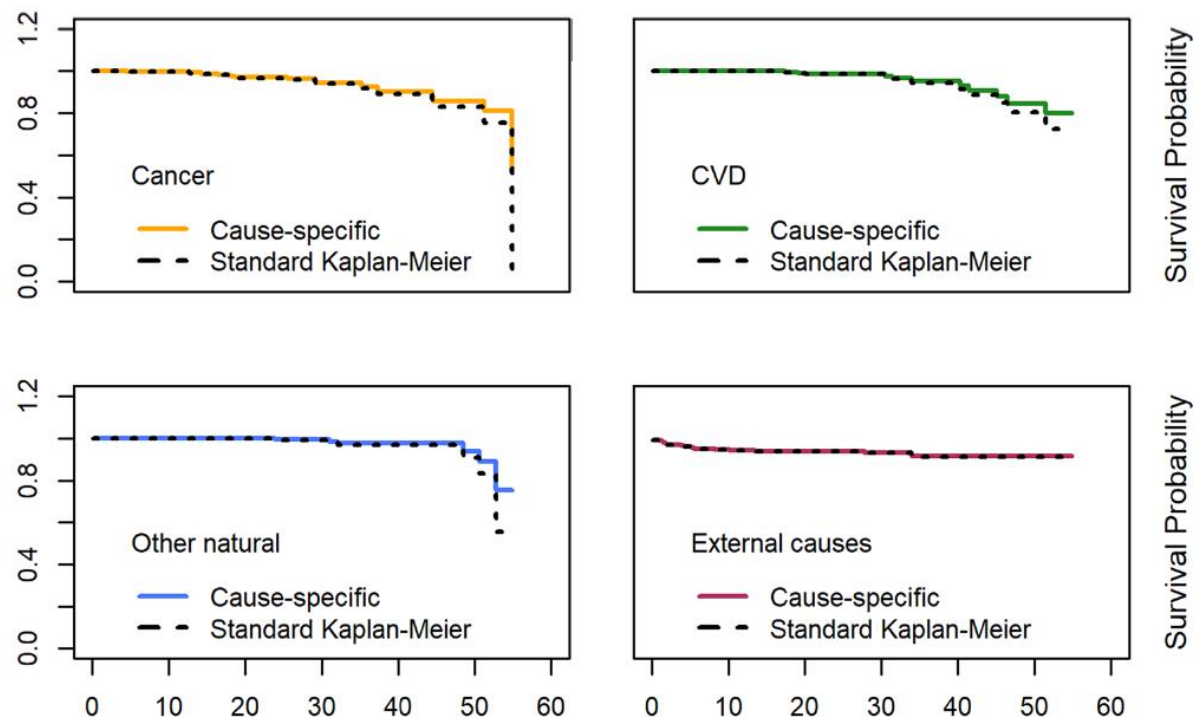


Example 1: Space radiation and mortality



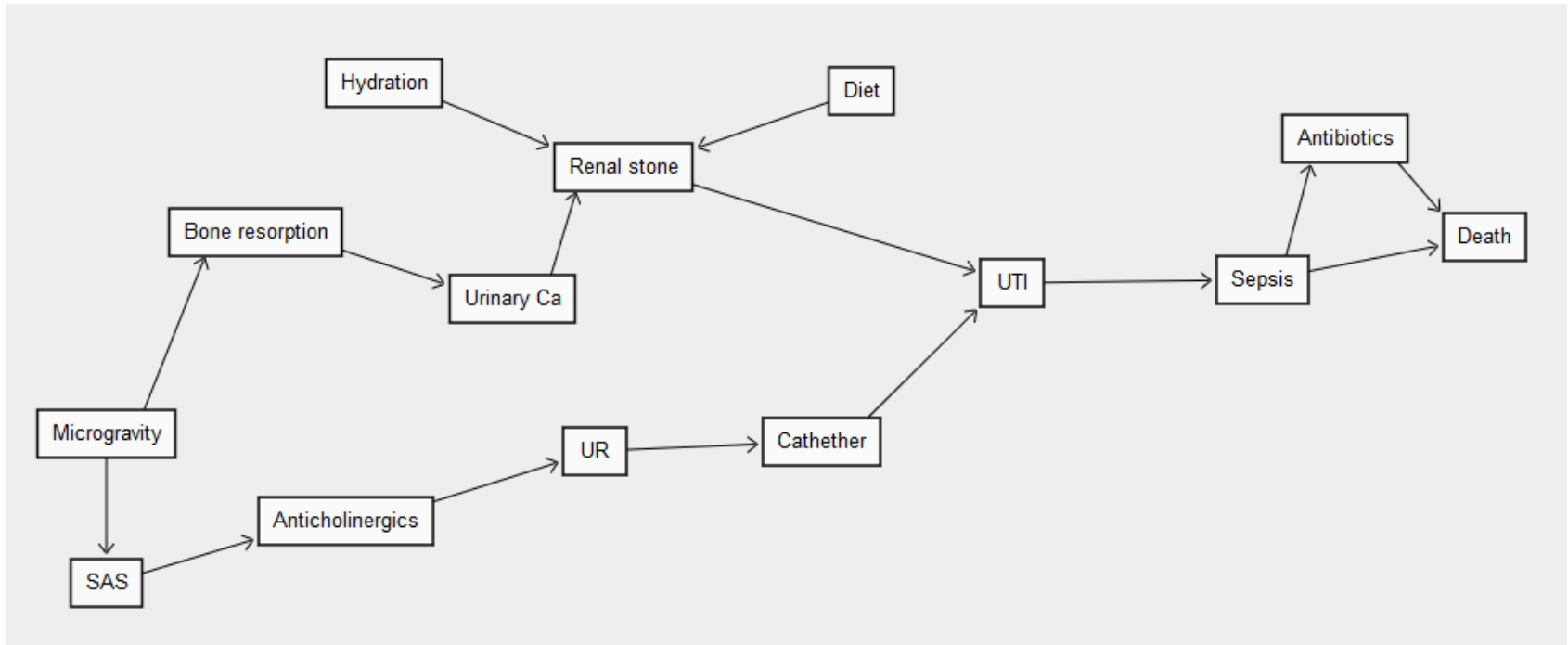
Alternative use of statistics: interference in survival curves.

- Divergence between standard Kaplan-Meier survival curves and competing-risk curves would be evidence of statistical dependence between causes of death





Example 2: DAGs / BN for spaceflight risks





Summary:

- NASA datasets are small, leading to low statistical power
- Non-random assignment to exposures creates problems in inference
- Interpretation of results becomes challenging
- NHST leads us to focus on avoidance of error rather than learning from data
- Scientists should think deeply about causal issues using causal assumptions and modern causal inference methods
- DAGs should be used to articulate hypothesized causal relationships
- Alternative hypothesis formulation and alternative analytic approaches should be employed as needed



Thank you!

Rob Reynolds, MS MPH PhD
rreynolds@mortalityresearch.com
[linkedin.com/in/robertreynoldsphd](https://www.linkedin.com/in/robertreynoldsphd)