Stop the abuse: A plea for a more principled approach to the analysis of adverse events

Kaspar Rufibach (F. Hoffmann-La Roche, Basel)

Tim Friede (University Medical Center Göttingen)

On behalf of the SAVVY working group

Disclaimer

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to Drug Information Association, Inc. ("DIA"), its directors, officers, employees, volunteers, members, chapters, councils, Special Interest Area Communities or affiliates, or any organisation with which the presenter is employed or affiliated.

How do we typically analyze «safety» in a clinical trial?

Typical reporting of AEs in a clinical trial

Category	All Adverse Events	
	Obinutuzumab Group (N=595)	Rituximab Group (N=597)
Infection*	460 (77.3)	418 (70.0)
Neutropenia	301 (50.6)	269 (45.1)
Infusion-related event†		
Any event	406 (68.2)	349 (58.5)
Antibody-related event	353 (59.3)	292 (48.9)
Tumor lysis syndrome	6 (1.0)	3 (0.5)
Cardiac event‡	78 (13.1)	58 (9.7)
Thrombocytopenia	68 (11.4)	45 (7.5)
Second neoplasm§	43 (7.2)	30 (5.0)
Nonmelanoma skin cancer	18 (3.0)	14 (2.3)
Hematologic event¶	6 (1.0)	0
Other	22 (3.7)	18 (3.0)
Myelodysplastic syndrome	2 (0.3)	0
Gastrointestinal perforation	4 (0.7)	3 (0.5)
Hemorrhagic event	57 (9.6)	62 (10.4)

N Engl J Med 2017;377:1331-44.

DOI: 10.1056/NEJMoa1614598

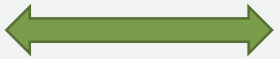
What do these proportions estimate?

- Incidence proportion for a group of patients in interval from 0 to t:

$$\widehat{IP}(t) = \frac{Number\ of\ patients\ with\ AE\ in\ [0,t]}{Total\ number\ of\ patients}.$$

- Estimates probability of AE happening in [0, t] and that **this AE is observed**.
- «this AE is observed»: implies that if we have (administratively) censored patients in our dataset then $\widehat{IP}(t)$ **underestimates true AE probability**.
- Interpretable if all patients have ~ same follow-up.

What do we want to report with safety analyses?

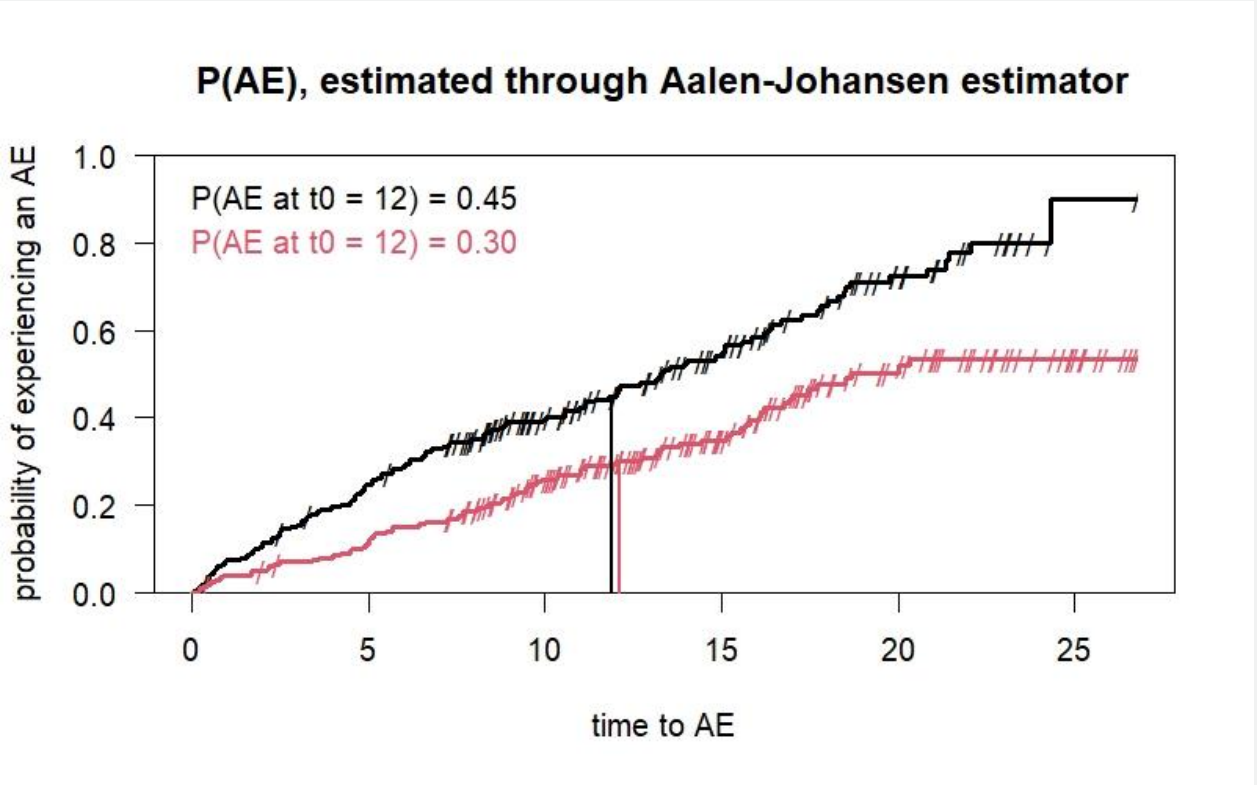
- Signal detection  assessment of AE risk.
- Here: assess AE risk for selected AEs.
- Probability of experiencing an AE, P(AE): **estimand**.

Challenges for assessment of AE risk

- (Administrative) **censoring**: patients not observed on entire [0, t]. Randomized «too late».
- **Competing events**: Clinical events that preclude the occurrence of AE.
  - Death without prior AE.
  - Treatment discontinuation leading to end of AE recording.
  - «Simply censoring» competing events gives biased estimates of probabilities.
- What is the **process that generates the events** (underlying hazard)?

Potential estimators of P(AE)

- Consider «time to AE» → methods from survival analysis.



Can we simply use 1 – Kaplan – Meier?

- For time-to-event endpoints (OS, PFS) we typically use Kaplan-Meier estimates → properly accounts for censoring.
- Do these work also for time to AE?
- Yes – in absence of competing events.
- Competing events: 1 – KM **overestimates** true P(AE).

Potential estimators of AE risk

	Accounts for censoring	Accounts for competing events	Makes no constant hazard assumption
Incidence proportion	**No**	Yes	Yes
1 – Kaplan-Meier	Yes	**No**	Yes
Aalen-Johansen estimator	Yes	Yes	Yes

- **Aalen-Johansen**:
  - Only estimator that accounts for censoring, competing events, and does not make assumption about underlying hazard.
  - Advocated for in statistical literature for decades.
  - «Gold standard». Available in any standard software package.

How large can the bias become in a clinical trial?

SAVVY = academia + pharma

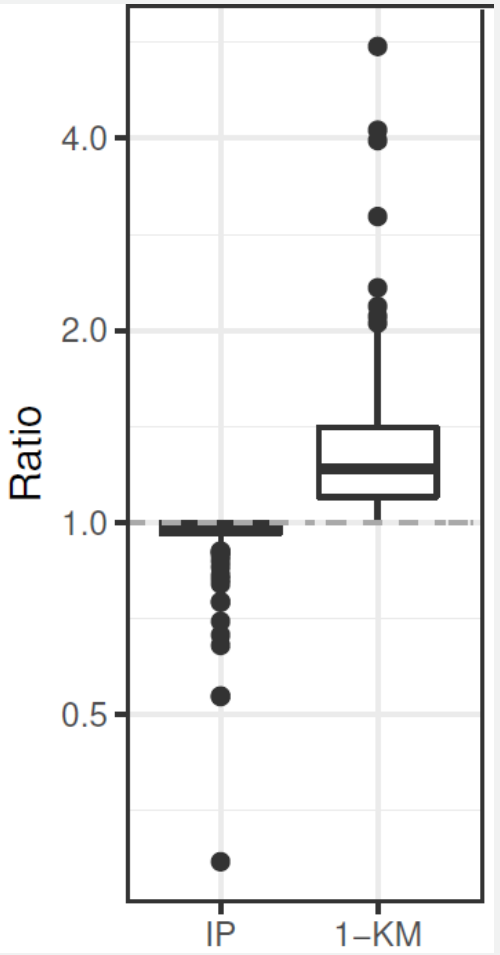
- **S**urvival analysis for **A**d**V**erse events with **V**ar**Y**ing follow-up times (SAVVY):
  - **Collaborative** effort from academia and pharma.
  - Goal: improve analyses of AEs in clinical trials through use **survival techniques** that account for varying follow-up times, censoring and competing events.
- Organizational setup:
  - **Centrally developed** R / SAS macros sent to organizations.
  - Summaries computed on clinical trial datasets within organizations → **raw data never left organization**.
  - Summaries centrally meta-analyzed.

How large can the bias become in a clinical trial?

- Dataset:
  - **Ten** organizations (9 pharma + 1 academic) provided **17 trials** including **186** types of AEs.
  - Trials included between 200 and 7171 patients (median: 443).
- «Gold standard»: Aalen-Johansen estimator of P(AE).
- Maximum follow-up time: compute P(AE) at the last observed AE time.

How large can the bias become in a clinical trial? Incidence proportion (IP)

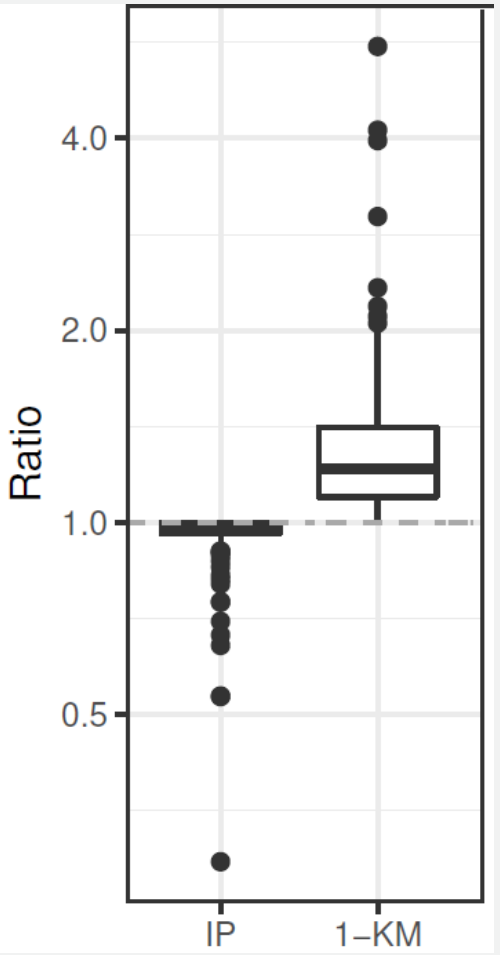
- Every observation in boxplot corresponds to ratio of IP to gold standard for a type of AE.
- Ratio = 1: IP gives same estimate as gold standard.
- **Underestimation of P(AE) up to factor THREE!**
- Overall reasonable performance.
- IP accounts for CEs but not censoring → datasets have many CEs → little censoring.



How large can the bias become in a clinical trial?

I – Kaplan-Meier (I - KM)

- Every observation in boxplot corresponds to ratio of (1 – KM) to gold standard for a type of AE.
- Ratio = 1: (1 – KM) gives same estimate as gold standard.
- **Overestimation of P(AE) up to factor FIVE!**
- (1 – KM) accounts for censoring but not CEs.



Does it impact decisions?

- SmPC frequency categories: very rare (< 0.01%), rare (< 0.1%), uncommon (< 1%), common (< 10%), very common (≥10%).
- Compare frequency category from IP or (1 – KM) to those from gold standard.

Does it impact decisions?

		gold-standard Aalen-Johansen				
		very rare	rare	uncommon	common	very common
incidence proportion	very rare	**6**				
	rare		**0**			
	uncommon			**6**		
	common				**86**	2
	very common					**86**
1-Kaplan-Meier	very rare	**6**				
	rare		**0**			
	uncommon			**4**		
	common			2	**72**	
	very common				14	**88**

- **Potential impact on labeling!**

How about estimation of relative effects?

- Categorization according to German Institute for Quality and Efficiency in Health Care (IQWiG).

		HR Cox for AE			
		(0) no effect	(a) minor	(b) considerable	(c) major
RR gold-standard Aalen-Johansen	(0) no effect	**42**	3	3	1
	(a) minor	9	**2**	1	
	(b) considerable	4	1	**3**	2
	(c) major	2		4	**17**

- **Potential impact on labeling!**

Conclusions

- Start with the **scientific question**: Signal detection? Assessment of AE risk?
- Respect **data structure**: censoring? Competing events?
- Estimand!
- **Select appropriate estimator**.
- Estimation of P(AE) in one cohort: Incidence proportion and (1 – KM) **biased** in presence of censoring or competing events.
- Comparing AE risk between two arms in RCT: differences between estimators become more emphasized.
- Template for:
  - Academia – pharma partnership.
  - Pool results of 17 RCTs without need for individual patient data sharing.

SAVVY: next steps and resources

- Estimate P(AE) in selected indications accounting for data structure.

- Statistical Analyis Plan:
  - Stegherr et al (2021) Biom J https://doi.org/10.1002/bimj.201900347
- Methods:
  - Preprint https://arxiv.org/abs/2001.05709
- Results:
  - One-sample case: https://arxiv.org/abs/2008.07883
  - Two-sample case: https://arxiv.org/abs/2008.07881
- Example R code (markdown file): https://numbersman77.github.io/AEprobs/SAVVY_AEprobs.html

SEE YOU NEXT YEAR

DIA EUROPE 2021
ADVANCING HEALTH PRIORITIES