# Eliminating Bias

Qudol

Complex data pipelines can lead to the introduction of systemic biases in research data. Qudol helps to reduce these biases by algorithmically linking data from any source with unlimited history.

One of the most well-known examples of data bias is shown to the right: the red dots represent a mapping of bullet holes in World War II aircraft. Abraham Wald from Columbia University's Statistical Research Group famously concluded that the density of bullet holes was inversely correlated with a need for more armor plating, recognizing that aircraft struck in the other areas simply didn't make it back home.
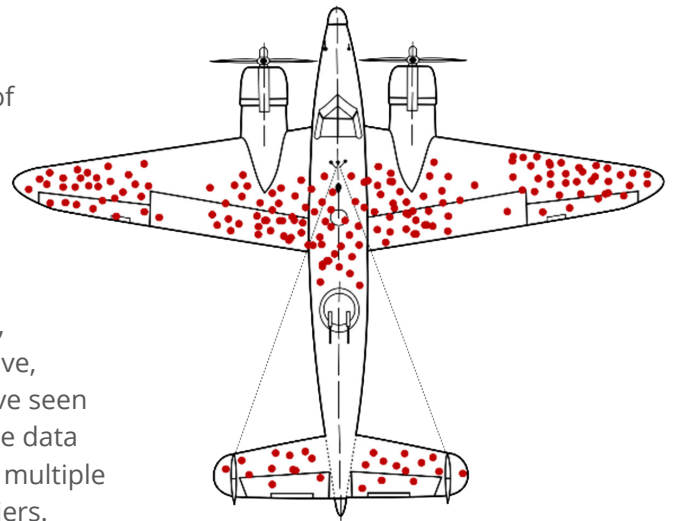
This is an example of **survivorship bias**, where failures are not taken into consideration when examining a dataset. Typically, biases such as these can lead to unintentionally non-representative, over-correlated and overly optimistic conclusions. Moreover, we've seen in practice that these issues often arise from the complexity of the data pipeline – the consequence of having multiple data sources from multiple data vendors with multiple sets of nonconforming unique identifiers.

Source: Wikipedia

Case Study: Teva Pharmaceutical

QUID*	AssetCode*	FromDate	Name	Trading Symbol
USA5555123	MSCIG-01234551	1996-12-31	TEVA PHARMACEUTICAL IND	TEVA.TA
USA5555123	MSCIG-01234562	2016-11-01	TEVA PHARMA IND ADR	TEVA

* Identifiers have been obfuscated for compliance. Names and dates may be fictitious, but meaningful.

Benchmark Index Event:
In 2016, MSCI replaced TEVA.TA (Tel Aviv) with TEVA (ADR/US) in the index because of limited liquidity.

Link to Announcement

This change to the MSCI dataset has the following implications for research and analysis:

- Fragmentation of TEVA Records: The rows corresponding to the two versions of Teva are not connected inside MSCI. If an analyst looks at what is happening within the index, she will only see the newer TEVA (ADR/US), and not the previous TEVA.TA that was dropped because of liquidity issues. This introduces **survivorship bias**.

- Loss of History: Looking at TEVA within the index, there are records going back to 2016. However, there's actually history going back to 1982 across multiple other data sources. The lack of linkages across history and across data sources introduces **exclusion bias**.

Qudol correctly assigns QUIDs to all variants, even with different SEDOL and CUSIP identifiers. The MSCI data does provides a hint in that the issuers are the same, but this is neither a definitive linkage (they could simply be different share classes), nor would this type of disambiguation be automatic or scalable. In this example, Qudol can connect these MSCI assets by cross-referencing BARRA data using the root ID.

The bottom line is that Qudol automatically resolves these issues within your dataset, for Teva and for countless other instances of corporate actions, benchmark changes, and other critical data events across time.

Qudol has over 25 years of successful partnerships with asset management teams. To book a demo and see more ways Security Master can improve accuracy and simplify your data operations, send us a message at info@securitymaster.com.