

Essay

Not peer-reviewed version

What 'Typical Insights' Might 'Core Statistical Analytical Techniques' Provide to Intelligence Analysts?

George Ellison *

Posted Date: 2 June 2026

doi: 10.20944/preprints202606.0055.v1

Keywords: statistics; intelligence analysis; big data



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Essay

What 'Typical Insights' Might 'Core Statistical Analytical Techniques' Provide to Intelligence Analysts?

George Ellison

Centre for Intelligence Studies at the University of Lancashire; gthellison@lancashire.ac.uk

Abstract

This article critically examines Jack Duffield's proposition that (all) intelligence analysts should become competent in "a core set of statistical analytical techniques" so as to address the "total information overload" they have experienced following the proliferation of 'Big Data'. It summarises the generic (technique-agnostic) and particular (technique-specific): analytical choices and decisions that analysts need to be competent to make when using each of the 11 techniques proposed; together with the parametric and non-parametric assumptions on which each of these techniques rely; the sources of non-systematic and systematic error that analysts using these techniques need to address; and the diagnostic measures that providers and consumers of findings generated by these techniques should use to assess any flaws or contingencies associated therewith, and thereby temper any associated inferential certainty and importance. When compared to baseline statistical competencies of non-specialist intelligence analysts, these summaries demonstrate the substantial additional training that Duffield's proposition would require. The article concludes that this may prove a big ask without an extended period of additional training, work-based experience and specialist supervision. In the absence thereof, underqualified intelligence analysts using such techniques would risk undermining intelligence analyses with multiple analytical and inferential mistakes.

Keywords: statistics; intelligence analysis; big data

Introduction

In Jack Duffield's audacious article ("Statistical Analytical Techniques for Intelligence Analysis" – published in the *RUSI Journal*; 2026), he argued that "a core set of [statistical] techniques... [might] form a strong baseline for a statistical analytical techniques primer for intelligence analysis"; and proposed 11 such techniques, for each of which he drafted examples of the "typical analytical insights they can provide".

A brief description of each of these "statistical analytical techniques" has been provided in Table 1 (overleaf), from which it is clear that the techniques proposed include few that will be familiar to many intelligence analysts – particularly the substantial proportion of whom only have "a relatively low mathematical baseline" (as Duffield himself acknowledges). While some knowledge and understanding of Linear Regression and Trend Plotting might be relatively widespread (not least as a result of the COVID-19 pandemic; Montgomery and Engelmann, 2020), the remainder comprise advanced, specialist techniques for which even statistically literate analysts might have limited knowledge and understanding, and little if any experience of their practical application.

For this reason, each of the "core" techniques, together with the "typical analytical insights" Duffield (2026) drafted, warrant more detailed explication and critical consideration given his suggestion that (all) intelligence analysts should be competent to apply these – and interpret their results – so as to cope with the "total information overload" he lays at the door of 'Big Data'.

Table 1. A brief methodological description of the 11 statistical techniques that Duffield (2026) proposed would “form a strong baseline for a statistical analytical techniques primer” for intelligence analysis. ¹

Technique		Methodological description
Linear Regression		Models continuous outcome/target variables as the weighted sum of covariates/‘predictor’ variables; generating coefficient estimates that minimize squared errors to accommodate the model’s adjusted variance, assuming independent, normally distributed residuals and linear relationships; and supports: causal inference of temporally consistent exposure-outcome relationships (subject to the mitigation of confounder/collider bias); and ‘predictive estimation’ of target variables through inter/extrapolation
Logistic Regression		Similar to Linear Regression except that the outcome/target variable is dichotomous and odds ratios are used to describe their relationships with covariates/‘predictor’ variables; but inference is limited to strength/direction/precision of causal relationships/‘predictive estimates’, not to linear/ordinal trends
K-Nearest Neighbours		A non-parametric algorithm that classifies observations by averaging the k closest data points in feature space; capturing local distributional structures (albeit sensitive to scaling and the choice of k); and can provide either grouping of categorical variables or estimates of continuous feature variables
K-Means Clustering		Partitions data into k groups by minimizing within-cluster distances to centroids; iteratively updating cluster assignments <i>and</i> centroids; revealing unmeasured group structure/pattern similarity; but dependent on the pre-specification of k , and the existence of/assumption that clusters are spherical
Analysis of Covariance (ANCOVA)		Similar to combining Linear Regression with Analysis of Variance (ANOVA) to compare group means while adjusting for continuous/categorical covariates; thereby assessing if adjusting for covariates alters coefficient estimates; and supports causal inference (subject to temporal consistency and bias mitigation) and ‘predictive estimation’ (assuming independent measurements and linear adjustment)
Trend Plotting		Visualizes changes in the value of variables over time (or within a sequence), often using smoothing or fitted lines to: identify growth/decline, cycles, turning points, and periodicities; guide modelling, hypothesis formation and ‘predictive’ estimation; albeit with limited mechanistic causal insight
Auto-Regressive Integrated Moving		A time-series model combining autocorrelation (AR), differencing (I), and moving-average smoothing (MA); capturing temporal dependencies, and using past measurements to (predictively) estimate future values; though, when data are stationary (or are made so), this technique can <i>also</i> support inference about persistence, cycles, and forecasting uncertainties
Bayesian Analysis		Combines prior beliefs with measured data using Bayes’ theorem to produce: <i>a posteriori</i> distributions that support probabilistic inference about parameters/hypotheses; with direct quantification of uncertainty; and dynamic learning as new data are added, are collected or become available
Monte Carlo Simulation		Uses repeated random sampling from specified distributions to estimate the range, probability, and sensitivity of outcomes in complex or uncertain systems; thereby providing probabilistic outcome distributions, risk assessments, and decision-reliability estimates
Cross-Validation		Assesses model generalisability by partitioning data into training and validation subsets; thereby evaluating estimative model performance on ‘unseen’ data; while being capable of detecting ‘overfitting’ and supporting model selection based on ‘out-of-sample’ accuracy measures

Bootstrapping	Resamples data with replacement to empirically ('predictively') estimate sampling distributions of finite datasets; providing standard errors/confidence intervals, and bias estimates even when strong parametric assumptions are absent; thereby assessing the level of uncertainty evident within (more modest, and) finite samples in the absence of more substantive sample sizes
---------------	--

¹Derived from the following canonical sources: Agresti (2015); Bishop (2006); Box *et al.* (2016); Casella and Berger (2002); Efron and Tibshirani (1993); Efron and Hastie (2016); Gelman *et al.* (2014); Goodfellow *et al.* (2016); Harrell (2015); Hastie *et al.* (2009) Hyndman and Athanasopoulos (2021); James *et al.* (2021); Kutner *et al.* (2005); McElreath (2020); Montgomery *et al.* (2021); Murphy (2012); Rubin (1987); Shalizi (2021); and Wasserman (2004; 2006).

Methods

To this end, and following on from a less technical critique of Duffield's argument (Ellison, 2026; reproduced in the Appendices), this article sets out to comprehensively evaluate its principal technical dependencies through detailed consideration of:

1. The conscious and deliberate, generic (technique-agnostic) and particular (technique-specific), choices and decisions analysts must (be competent to) make when using each of the 11 techniques – and the potential analytical and inferential consequences were any of these choices/decisions to be misjudged, or overlooked/made passively ('by default');
2. The generic (technique-agnostic) and particular (technique-specific) parametric and non-parametric assumptions that need apply when using each of the 11 techniques – and the potential analytical and inferential consequences were any of these assumptions to be violated;
3. The generic (technique-agnostic) and particular (technique-specific) sources of non-systematic error (imprecision) and systematic error (bias) that can affect the reliability, validity and interpretation of findings generated using each of the 11 techniques – and the analytical and inferential consequences were steps not to have been taken so as to avoid, attenuate, mitigate, or acknowledge (and accommodate) these sources of imprecision/bias (and their effects);
4. The diagnostic assurance measures that *producers* of 'statistical analyses'¹ – i.e. those involved in identifying/selecting/collating/generating, processing and analysing quantitative/enumerable datasets, such as those working within the UK's intelligence collection "disciplines" (MOD, 2023) – can, and should take to ensure, evidence and reassure that: the most appropriate analytical choices/decisions have been made (see 1, above); all necessary assumptions apply (see 2, above); and all potential sources of imprecision and bias have been: avoided or attenuated (where possible); their effects mitigated (again, where possible); or any residual bias acknowledged (as a potential influence on the findings obtained) and accommodated (in the level of uncertainty and confidence applied when interpreting these findings; see 3, above) – these being diagnostic measures that should have been implemented within the reported design and implementation of the analyses concerned (or may need to be retrospectively applied during translation), so that *consumers* of statistical analyses (such as those working within the intelligence analysis "specialisms"; MOD, 2023) are aware of any potential gaps, weaknesses, flaws and residual uncertainties in the findings generated by these analyses;
5. Potential misrepresentations evident in the "typical analytical insights" Duffield (2026) provided for each of these 11 techniques (see Table 1, above), and potential misinterpretation of the inferences these "insights" might support – subject to: the proficiency with which each of the 11 techniques have been applied (see 1 and 2, above); the extent to which any sources of imprecision and bias have been addressed or acknowledged (and accommodated; see 3, above); and the diagnostic measures that producers of statistical analyses took to ensure, assure (and evidence) that their analyses were both rigorous and robust (see 4, above); and

6. Whether the knowledge and skills required to accommodate each of the considerations summarised in 1 through 5 (above) might plausibly be accessible to analysts operating from what Duffield (2026) acknowledged was “a relatively low mathematical baseline”.

This article addresses each of these six considerations in turn, and will conclude by evaluating both: the practicability of Duffield’s (2026) proposition – that (all) intelligence analysts should become competent in “a core set of statistical analytical techniques” – without recourse to an extended period of training and/or specialist supervision; and the risk that inadequate training/expertise in such techniques might vitiate intelligence analyses with multiple, and potentially damaging, analytical and inferential mistakes.

Results

1. What Choices/Decisions Must Analysts Make When Using the 11 Techniques?

A non-exhaustive list of the generic (technique-agnostic) and particular (technique-specific) choices and decisions analysts must (be competent to) make when using any of the 11 techniques Duffield (2026) proposed, have been summarized in Table A1.1-A1.3 (see Appendices). In the main, the choices/decisions summarised in Table A1.1 and A1.2 focus predominantly on design- and dataset-related considerations, each of which should themselves be determined in line with the ‘analytical objective’ – i.e. with the choice/decision made as to the intended application of the analyses concerned, be these for descriptive, ‘predictive estimation’, optimisation or causal inference purposes. All of these overarching analytical objectives benefit from datasets that offer a representative sample of all ‘cases’ (i.e. the ‘population’) to which their analytical findings can then be applied – be this a ‘population’ of entities, events, processes, or characteristics thereof. However, for ‘predictive estimation’ and optimisation objectives,² the principal benefit of representative datasets is that the ‘predictive estimation’ and optimisation algorithms generated will be broadly generalisable to all other similarly representative datasets elsewhere. However, where only *unrepresentative* datasets are attainable/available, useful ‘predictive estimation’ and optimisation algorithms can still be usefully generated, though *only* for subsequent application to *similarly unrepresentative* datasets – including the original dataset concerned, or any other comparable datasets that have been sampled/selected in a similarly unrepresentative fashion – such that they contain sufficiently similar distributional properties on which the algorithms generated can operate faithfully.

The remaining choices and decisions within Tables A1.1 and A1.2 fall into those concerned with: the size of the sample required to support the analyses and insights intended, and the levels of precision required; what specific variables the dataset needs to contain; how these variables need to be prepared/processed prior to analysis; and what diagnostic measures/settings are required to optimise the accuracy, precision and interpretability of the resulting analytical findings (and any insights that might be inferred therefrom). In this regard, irrespective of the analytic objective intended, it is crucial that:

- the number of ‘cases’ (i.e. ‘population’ members) on which measurements are available is sufficient to both: support the analytical techniques applied; and reduce the risk of the findings being subject to chance effects (i.e. non-systematic error);
- measurements are available on all cases for all variables pertinent to (and required for) the sample- and variable-dependent analyses intended (and that any under-/over-represented cases can be dealt with, where necessary/appropriate, using robust weighting variables; and any missing values can be accurately imputed – and not least when missing cases/values are not missing [completely] at random, and therefore risk undermining the representativeness of the dataset); and
- these variables include not only the specified ‘target’ variable (for ‘predictive estimation’ and optimisation objectives) or ‘outcome’ variable (for description and causal inference objectives), but also a sufficiently varied array of covariates/predictor variables so as to optimise the

diversity of statistical information required to: strengthen the accuracy and precision of the estimated dataset features (for 'predictive estimation' and optimisation objectives); and adjust for potential confounders (for causal inference purposes).

Deciding how many measurements, and of which variables, are necessary to collect (or need to be present within the secondary dataset selected), can prove an uncertain art and statistically challenging to operationalise – particularly when there is limited time and resource available to generate datasets that meet/exceed the 'optimum' size and diversity desired; or when the secondary datasets available do not contain either the number of measurements or range of variables desired. In such instances – which are commonplace in real world settings – the analysts concerned will need substantial skill to balance the benefits of sample size over diversity (notwithstanding representativeness) when compiling or selecting the datasets they use. They also need to be able to: judge when the size and/or diversity of the dataset required for their intended analytical objective is not achievable/available; and assess whether it is better to step back from attempting to achieve their desired/intended analytical objective using datasets that are patently incompatible with doing so confidently (or, at least, without a substantive risk of excessive imprecision and/or bias).

Meanwhile, even those datasets that *do* contain sufficient measurements from sufficiently diverse variables may require these data to be carefully pre-processed so as to: optimise the variance-determined statistical information available for use by the analytical technique(s) concerned; and facilitate the interpretation (and communication) of any findings generated. This includes not only procedures to address missing values (as mentioned earlier), but also – subject to the assumptions required by the analytical technique applied (see Section 2, below) – any necessary transformation of the raw data into meaningfully ordinal or categorical variables, and to normalise and/or linearise the distribution of their measurements/variances and/or model residuals. Overlooking or misjudging these considerations can lead to imprecise, inaccurate, unreliable and invalid findings; or findings that are unnecessarily challenging to interpret or explain to non-specialist *consumers* of the insights produced. Again, these decisions require great care and thoughtfulness so as to: not simply ensure that the sampling, precision, linearity and/or distributional requirements are met; but also to hold fast to the intended/desired analytical objective and the inferential needs of their customers.

Finally, aligning the diagnostic measures selected for testing and evaluating the integrity of the analyses³ with the size and diversity of the dataset available, the analytical technique selected, *and* the intended analytical objective, is *also* an important consideration. This is particularly the case when weaknesses in *any* of these analytical components, or a lack of firm empirical knowledge and/or theoretical understanding, increase the extent of analytical uncertainty and thereby reduce analytical confidence. Such measures need to be fit for purpose if they are to challenge the integrity of the analyses and offer confidence that their findings can survive such testing/evaluation, and can therefore be considered robust.

These (technical) considerations come more to the fore in Table A1.2 (summarising the technique-agnostic choices/decisions that are particular to Duffield's (2026) 11 proposed statistical analytical techniques) and Table A1.3 (summarising the technique-specific choices/decisions that are particular to smaller subsets – and occasionally just one – of these 11 techniques). Together these illustrate the depth of technical knowledge and expertise required to competently apply each and every one of these 11 statistical analytical techniques – knowledge and expertise that goes some way beyond the more generic quantitative analytical competencies summarised in Table A1.1.

2. What Parametric/Non-Parametric Assumptions Need Apply When Using the 11 Techniques?

The generic (technique-agnostic) and particular (technique-specific), parametric and non-parametric assumptions that should apply when using each of Duffield's (2026) 11 proposed techniques have been summarised in Table A2.1 and A2.2, respectively (see Appendices). The first of these, for the most part, comprise assumptions that commonly apply to a far wider range of statistical techniques – these being those related not only to the analytical objective (and the constraints this can impose on the objective function/loss criterion modelled), but also to:

- the representativeness/external validity of the dataset used;
- the reliability and internal validity of the measurements taken/available on each of the variables therein;
- the appropriateness of the data type, scaling and transformation(s) adopted for/applied to these variables – particularly as required for the statistical technique concerned;
- the independence of measurements/observations made for each of the dataset's/sample's individual 'cases'/population members;
- the absence (or appropriate treatment of) extreme ('outlier') measurements/observations;
- the appropriateness of the size of the dataset/sample, and any necessary randomness assumptions (where relevant to re-/sub-sampling procedures germane to the statistical technique applied); and
- a sound theoretical understanding of the temporal sequence of variables – particularly for modelling time-dependent relationships/associations (such as causal pathways).

As such, these assumptions add to many of the dataset- and data-related choices/decisions that analysts need to (be competent to) make when using *any* (or at least *most*) statistical techniques – as summarised in Section 1 (above), and in Tables A1.1 and A1.2. As such, these comprise a broader corpus of considerations and obligations that analysts need to understand and apply whenever they attempt (*m/any*) statistical analytical techniques.

However, the more particular (technique-specific) assumptions that need to apply (predominantly, though not necessarily exclusively) to each of Duffield's (2026) 11 proposed techniques – as summarised in Table A2.2 – are substantively different to the parallel choices/decisions contained in Table A1.3. Though *some* of these assumptions determine what choices/decisions analysts must make when designing, implementing and interpreting the results generated by these techniques, most relate to more granular technical issues that distinguish each technique as somewhat unique – and again speak to the depth of specialist knowledge, understanding and expertise that analysts require when applying these.

3. What Are the Potential Sources of Imprecision and Bias When Using the 11 Techniques?

The potential sources of non-systematic and systematic error (i.e. imprecision and bias) summarised in Table A3.1 and A3.2 (see Appendices) can occur at each and every stage of any (statistical) analytical process – from design, sampling and measurement through to analysis, interpretation and implementation. They are also evident in any flaws in the design and application of the specific (statistical) analytical process concerned, *and* in any knock-on consequences of the interpretations made and inferences drawn from their apparent (yet misinterpreted) findings. Arguably, each of these flaws primarily occurs as a result of mistaken choices/decisions or violated (overlooked/untested) assumptions amongst those summarised in Tables A1.1 through A2.2 (see Section 1 and 2, above). This is because imprecision and bias are in no small part the result of poor (statistical) analytical practice, and as such are a prominent feature of the problems that can arise when analysts with inadequate knowledge, understanding and expertise attempt to apply (and interpret findings generated by) these techniques without recourse to specialist guidance, advice, supervision and support.

Yet the very existence, importance and potential consequences of a good many of the sources of imprecision and bias summarised in Table A3.1 – and even more so for those summarised in Table A3.2 – remain somewhat contentious and contested amongst qualified and experienced statistical professionals. As a result, many are widely overlooked or ignored in what passes for contemporary professional statistical practice (or the canonical works consulted when preparing Table 1, above). In part this reflects the differing pace of discovery and change in the fields of theoretical (or 'classical') statistics, and the more applied and contextually cognisant discipline of 'biostatistics' – in which the latter faces more frequent opportunities for identifying unacknowledged flaws in accepted techniques through their use in novel data-centric problems, contexts and expectations. This has two consequences for the wider application of statistical analytical techniques by intelligence analysts:

- First, that many ‘standard’ techniques may prove less applicable to the data and datasets that are available, and of most interest, to intelligence practitioners and their customers; and
- Second, that these data/datasets are likely to pose novel statistical challenges to many of the more commonly used statistical analytical techniques.

Examples of the former include the greater number and importance of predominantly unprecedented, singular phenomena/events amongst those that constitute “priority intelligence requirements” (PIRs; MOD, 2023) – phenomena/events *for* which, and *on* which, there is often a dearth of comparable past measurement/observation amenable to statistical analysis.⁴ As for examples of the second, in the absence of opportunities for controlled experimentation (which may nonetheless exist within some disciplines and domains, such as OSINT and CEMA; MOD 2023; 2024), the mechanistic understanding required for causal inference and foresight – as opposed to the estimation of unmeasured/unobserved future dataset features achieved using ‘predictive analytics’ – must rely entirely on careful consideration of temporality, and relentless efforts to address confounding bias while avoiding conditioning on variables acting as ‘colliders’ (Ellison, 2023). This is because *only* variables that occur (or crystallise in the form and value as, and when, measured/observed) *after* other variables can be plausibly considered their potential consequences; and *only* variables that occur (or crystallise) *before* other variables can be plausibly considered their potential causes.

Added to these is the inherent *variability* (and therefore intrinsic, if ultimately finite, *unpredictability*) of all life forms as a consequence of the intra- and inter-generational conservation of phenotypic and genotypic variability/diversity within changeable ecosystems that is central to the tenets of Darwinian natural selection. This is a relevant consideration in such contexts because these include the ‘extended phenotypes’ of human beings – as evident in the substantial diversity and ongoing diversification in perceptions, beliefs, ideas, understanding, explanations, interpretations, behaviours, cultures and artefacts – that each contribute ostensibly limitless sources of both non-systematic and systematic error (imprecision *and* bias) to *all* human-related intelligence concerns.

4. What Diagnostic Measures Should Producers of Statistical Analyses Use to Safeguard Their Methods?

As a safeguard against inappropriate choices/decisions, violated assumptions, imprecision and bias – and to evidence rigorous and robust statistical analytical practice – the routine application of a range of generic and specific diagnostic measures capable of identifying methodological concerns is gradually becoming inculcated into what passes for applied statistical doctrine (e.g. Lang and Altman, 2015; Ioannidis, 2017; 2019; Christensen et al., 2023; Hardwicke et al., 2023; Kim et al., 2024). As before, some of these diagnostic measures have broad utility irrespective of the statistical technique applied, while others predominately relate to less common features or peculiarities of the specific technique(s) involved. Non-exhaustive summaries of each – as relevant to the 11 techniques examined in this article – have been provided in Tables A4.1 and A4.2 (see Appendices). In the main, both of these Tables include diagnostic measures that help identify whether the analysts concerned have:

- made appropriate choices and decisions when specifying/selecting their analytical objectives, statistical analytical designs, sampling frames/secondary datasets, measurement/observation protocols, analytical models and inferential interpretations – as summarised in Table A1.1-A1.3, and described in Section 1 (above);
- designed and conducted their analyses using analytical objectives, datasets, analytical models and inferences that comply with the assumptions required of all (or applicable principally to) specific statistical techniques – as summarised in Table A2.1-A2.2, and described in Section 2, above; and
- applied great care to each of the (design-, sampling/selection-, measurement/observation-, analysis- and interpretation-related) steps involved so as to avoid, attenuate, mitigate, or acknowledge (and accommodate) all relevant sources of imprecision/bias (and thereby their consequences/effects) – as summarised in Table A3.1-A3.2, and described in Section 3 (above).

As a result, there are few surprises amongst either the generic or specific diagnostic measures listed in these Tables given they relate to each of the choices/decisions, assumptions and sources of imprecision and bias that have already been described in preceding Sections of this article. What is nonetheless surprising are the number of *additional* procedures that competent analysts will need to have mastered to ensure their statistical analyses can be relied upon within the context of strategically, operationally and tactically critical intelligence analysis, fusion and assessment. The point being that far from it being sufficient for intelligence analysts to master the “basic”, “baseline” or “core” principles and techniques required to confidently analyse substantive quantitative/enumerable datasets (Duffield, 2026), they *also* need to know enough to: make good choices/decisions; pay attention to the assumptions on which the techniques they use depend; address (or acknowledge and accommodate) all potential sources of imprecision and bias; *and* critically evaluate their own analyses using diagnostic measures capable of identifying critical weaknesses, flaws, conditionalities or contingencies. This means that with statistical analytical techniques the burden of expertise is high, and – given: ongoing advances in methodological understanding; and in the technologies available for exposing (and addressing) *some* longstanding/newly discovered analytical flaws/weaknesses; and the diversification of datasets and questions to which these techniques are applied – all statistical analysts (and all statistically competent intelligence analysts) need to keep up with ongoing developments in professional practice.

5. What Alternative Interpretations/Inferences Might Each of Duffield's 11 “Typical Analytical Insights” Support?

Taken together, the four preceding considerations (as examined in Sections 1-4, above) set a high bar for the statistical expertise expected of intelligence analysts were they to take on additional *statistical* responsibilities – for collecting/sourcing, processing and analysing large quantitative (and enumerable) datasets; and applying robust diagnostic measures so as to evaluate and quality-assure their analyses – *alongside* their principal function as single/multi/all-source *intelligence* analysts. Added to this will be the skills required to – succinctly-yet-accessibly – explain to customers (including: juniors; peers; more senior colleagues; commanders; decision-makers... few of whom will have specialist statistical expertise): what inferences they have been able to generate from their analyses; and what evidence they have that any resulting insight/foresight might be considered both reliable *and* robust.

Neither of these skillsets are for the faint-hearted, nor for those who lack either the time or commitment required to: apply such techniques rigorously; and translate them assiduously for non-specialist audiences. Yet the confidence that comes from experience of using statistical methods – as with any other advanced/specialised technique – can substantively attenuate the scale of the work involved. This is because ‘statistical analyses’^{1,4} are “not looking for perfection, but for advantage” (as the protagonist, George Smiley, described Steed-Asprey’s ‘double cross’ [XX] system in Le Carre’s celebrated 1974 novel, *Tinker, Tailor, Soldier, Spy*).

In statistical lore, a comparable and equally memorable aphorism can be found in the wise words of another George (George Box FRS – a much admired and distinguished statistician) who, at around the same time as Le Carre, wrote that “All [statistical] models are wrong, but some are useful” (Box et al., 1979). What both these Georges mean is that neither of these practices (be they: deception operations, for Smiley; or statistical inference, for Box) nor the techniques they use (be they: the ‘double cross’ [XX] system, for Smiley; or quantitative modelling, for Box) need to be mistake-, imprecision- or bias-free to be of *potential* tangible value/benefit to the problem at hand. Moreover, in an earlier article, Box (1976) had already set out what his (subsequently published) aphorism might mean for “scientists” – and by extension, for intelligence analysts and statistical analysts alike – namely that:

“Since all [statistical] models are wrong the scientist [statistician, or analyst] must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.”

The relevance of this to Duffield's (2026) proposition is that analysts who are committed to learning how to competently apply statistical analytical techniques will also learn, from practice and experience: which of the choices/decisions, assumptions, and sources of imprecision and bias pose the *most important risks* to the reliability, validity and substantive utility of their statistical analytical techniques; *and* which might only pose trivial concerns to the insights these techniques can provide.

The reason for labouring this point at this stage in the present article is to frame the scope of the critique that follows of the "typical analytical insights" Duffield (2026) drafted for each of his 11 proposed "statistical analytical techniques". Since the detail provided in these "insights" was necessarily brief (given their intent, and the finite word count available to him within the confines of an academic journal article) they do not include the more detailed references to (m)any of the methodological and technical considerations explored in Sections 1 through 4 (above) – information that might ordinarily be included in summaries of such findings within contemporary statistical publications/reports. As such, the critique that follows pays less heed to the *absence* of this more detailed methodological information, and instead assesses whether these (necessarily brief) examples of the "typical analytical insights", that each of the 11 techniques are purported to able to provide, might:

- *misrepresent* what each technique involves or what findings it can provide; and thereby lead to the
- *misinterpretation* of what might properly be inferred from a more cautious summary of their findings.

These assessments, together with a more circumspect 'alternative description' of each "typical analytical insight", have been summarised in Table 2, overleaf; and are described in more detail in Sections 5.1 and 5.2, below.

Table 2. (part 1 of 3). An assessment of the potential methodological misrepresentation and inferential misinterpretation of the “typical analytical insights” Duffield (2026) drafted for each of the first 4 of the 11 “statistical analytical techniques” he proposed intelligence analysts should be competent to apply and interpret; together with alternative descriptions intended to attenuate the risk of misrepresentation or misinterpretation.

“Technique”	“Theme”	“Typical Analytical Insight”	Potential methodological misrepresentation	Potential inferential misinterpretation	Alternative description
Linear Regression	Regression	<i>“After controlling for other variables, increases in X are consistently associated with proportional increases in Y, suggesting X is a meaningful driver rather than background noise.”</i>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● Reference to “controlling for other variables” is appropriate, but only were the residual association between X and Y (after controlling for these <i>specific</i> “other variables”) be the intended/desired estimand. 	<ul style="list-style-type: none"> ● The phrase “meaningful driver” implies that the association between X and Y can be confidently interpreted as consistent with X causing Y. ● This association is also consistent with: chance; reverse causality; collider bias; and unmeasured and/or residual confounding. 	<p><i>“X and Y are positively associated with one another after adjusting for all/all preceding* measured covariates.”</i></p> <p>*The inferential interpret-ability/-ion of the association depends upon which covariates were adjusted for, hence this warrants clarification.</p>
Logistic Regression	Regression	<i>“The probability of event A rises sharply once factors X and Y are present, indicating a threshold effect that distinguishes high-risk cases from the baseline.”</i>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● Use of the term “probability” to mean ‘likelihood’, ‘rate’, ‘risk’ or ‘odds’ is potentially confusing. ● A “threshold effect” would require additional evidence of a more or less ‘stable state’ prior to an ‘inception point’ of some sort. 	<ul style="list-style-type: none"> ● The phrase “threshold effect” risks misinterpretation as definitively causal (as in ‘cause and effect’). ● The increase in the odds of A in the presence of X and Y is also consistent with: chance; reverse causality; mathematical coupling; collider bias; and unmeasured and/or residual confounding. 	<p><i>“The odds that A occurs is positively associated with the presence of (both)* X and Y.”</i></p> <p>*Subject to clarification of the “typical analytical insight” as originally drafted</p>

K-Nearest Neighbours	Classification	<p><i>“This actor’s recent behaviour most closely resembles that of a small subset of previously observed cases, which historically went on to exhibit outcome B.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● The use of the term “most closely resembles” is appropriate but its utility depends on which “recent behaviour[s]” were involved. ● The “small subset” of cases which exhibited “outcome B” suggests ‘conditioning on the outcome’ may have generated this association. 	<ul style="list-style-type: none"> ● The phrase “historically went on to” risks misinterpretation as definitively causal. ● The elevated risk of “going on to exhibit outcome B” given the classification of the “actor’s recent behaviour” will also be consistent with: chance; reverse causality; collider bias; and unmeasured and/or residual confounding. 	<p><i>“(On the basis of the recent behavioural features measured)* This actor is most closely associated with past cases that later exhibited outcome B; but this actor’s risk of outcome B will depend upon the continuing presence of any other necessary factors/features.”</i></p> <p>*Subject to the limited scope of the behavioural features on which K-means classification was performed.</p>
K-Means Clustering	Classification	<p><i>“Observed entities naturally separate into four distinct behavioural groupings, each with internally consistent patterns and materially different risk profiles.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● This analytical technique (alone) is not able to determine: how the groupings observed arise; the nature/extent of any “internally consistent patterns”; or group-related differences in “risk profiles”. 	<ul style="list-style-type: none"> ● The use of “naturally separates into”, “materially different” and “risk profiles” infer a level of mechanistic understanding, substantive knowledge, and/or (additional) empirical analysis that lies beyond what this technique (on its own) can determine. 	<p><i>“(On the basis of the behavioural features measured)* Four distinct and internally consistent clusters were identified (and separate analyses indicate these clusters have different risk profiles for X [to be specified])”</i></p> <p>*Subject to the limited scope of the behavioural features on which K-means classification was performed.</p>

Table 2. (part 2 of 3). An assessment of the potential methodological misrepresentation, and inferential misinterpretation of the “typical analytical insights” Duffield (2026) drafted for next 4 of the 11 “statistical analytical techniques” he proposed intelligence analysts should be competent to apply and interpret; together with alternative descriptions intended to attenuate the risk of misrepresentation or misinterpretation.

“Technique”	“Theme”	“Typical Analytical Insight”	Potential methodological misrepresentation	Potential inferential misinterpretation	Alternative description
-------------	---------	------------------------------	--	---	-------------------------

ANCOVA (Analysis of Covariance)	Causal Inference	<p><i>“Once baseline capability and environment are accounted for, the apparent gap between actors narrows significantly, indicating that much of the observed difference reflects starting conditions rather than divergent behaviour.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is <i>inappropriate</i> – ANCOVA is a ‘general linear model’ that can only identify associations between variables (but see below). ● Such associations can nonetheless support causal inference where they examine: focal relationships between preceding and subsequent variables; and <i>after</i> competently mitigating the risk of confounder and collider bias. 	<ul style="list-style-type: none"> ● The original “typical analytical insight”, as drafted, presents only limited potential for inferential misinterpretation. ● The use of the phrase “narrows significantly” implies that the technique routinely provides a statistical measure of the similarity between ‘actor types’ (which it does not) and that this measure achieved a probability considered “significant”. 	<p><i>“The extent of the difference observed between different types of actors is substantively reduced following adjustment for baseline capability and environment, suggesting that much of the difference observed is due to their different starting conditions rather than subsequent developments”*</i></p> <p>*This is somewhat vague without information on what “differences” between actors were examined.</p>
Trend Plotting	Time Series	<p><i>“The underlying trajectory shows a sustained upward movement over multiple periods, with short-term volatility masking a longer term structural change.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● The use of the phrase “structural change” may exceed the evidence (of ‘underlying mechanism’) this technique can provide – although this will ultimately depend upon what variables/features have been modelled in this analysis. 	<ul style="list-style-type: none"> ● There is a risk that the phrase “structural change” may encourage solely those inferences that focus on (structural) mechanistic features rather than also on circumstantial/contextual changes occurring in a similar fashion over time. 	<p><i>“Shorter-term fluctuations in the measurements of this feature mask a sustained, longer-term upward trajectory indicative of a possible change in the prevailing conditions or underlying mechanisms involved.”*</i></p> <p>*This is necessarily abstract in the absence of information on the feature, time-frames and context concerned.</p>

<p style="text-align: center;">ARIMA (Auto-Regressive Integrated Moving Average)</p>	<p style="text-align: center;">Time Series</p>	<p><i>“Assuming current dynamics persist, activity levels are likely to remain within a bounded range over the next three periods, with a non-trivial risk of a sharp deviation thereafter.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● The use of both “likelihood” and “risk” to describe equivalent properties of the temporal trend described risks implying the latter involved additional analyses (i.e. other than ARIMA). 	<ul style="list-style-type: none"> ● The inclusion of the phrase “assuming current dynamics persist” helps ensure the inference implied is interpreted as conditional on ‘all else being equal’ (i.e. <i>ceteris parabus</i>). 	<p><i>“Conditional on the fitted ARIMA dynamics, forecasts of measurable activity levels are likely lie within X%* intervals for three periods, but less likely to remain within these intervals thereafter.”</i></p> <p>*To better define “a bounded range”.</p>
<p style="text-align: center;">Bayesian Analysis</p>	<p style="text-align: center;">Probabilistic Modelling</p>	<p><i>“Given the new reporting, confidence in hypothesis A has increased substantially, while alternative explanations now carry materially lower probability.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● The use of the term “confidence” in this (analytical) setting is problematic given its use to signify ‘precision’ in statistics, and ‘evidentiary strength’ in intelligence analysis. 	<ul style="list-style-type: none"> ● The (mis)use of the term “confidence” as a synonym for likelihood/probability implies that <i>both</i> the likelihood <i>and</i> the precision of “hypothesis A” has increased; while <i>only</i> the likelihood of alternative hypotheses have declined. 	<p><i>“The a posteriori probability for hypothesis A increased following the availability of this new evidence, with the probabilities of mutually exclusive alternative hypotheses being correspondingly reduced.”</i></p>

Table 2. (part 3 of 3). An assessment of the potential methodological misrepresentation, and inferential misinterpretation of the “typical analytical insights” Duffield (2026) drafted for final 3 of the 11 “statistical analytical techniques” he proposed intelligence analysts should be competent to apply and interpret; together with alternative descriptions intended to attenuate the risk of misrepresentation or misinterpretation.

“Technique”	“Theme”	“Typical Analytical Insight”	Potential methodological misrepresentation	Potential inferential misinterpretation	Alternative description
-------------	---------	------------------------------	--	---	-------------------------

Monte Carlo Simulation	Probabilistic Modelling	<p><i>“Over thousands of simulations, four plausible scenarios emerged, including a sharp deterioration scenario, which occurs when both X and Y occur at a similar time.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● Inadequate detail on the number of emerging/evident scenarios (regardless of “plausibility”). ● Citing “thousands of simulations” (not unusual for this technique) is superfluous and misrepresents the assiduousness of its application. 	<ul style="list-style-type: none"> ● There is a modest risk of unintended/implicit causal inference from “when both X and Y occur” – particularly were (i) “at a similar time” to be taken for ‘just before’; or (ii) no subsidiary analyses to have been performed so as to explore all possible ‘variable-scenario co-occurrences’. 	<p><i>“(Under the assumed input distributions, at least)* four scenarios are evident, one of which involves a sharp deterioration that is associated with the co-occurrence of X and Y”</i></p> <p>*The original “typical analytical insight” would benefit from adding this disclaimer and clarifying how many scenarios are evident.</p>
Cross-Validation	Cross-Validation	<p><i>“The model’s predictive performance remains stable across unseen data, indicating that the identified patterns are likely to generalise rather than reflect overfitting.”</i></p>	<ul style="list-style-type: none"> ● The thematic labelling of this technique is appropriate. ● The only modest risk of misrepresentation lies in the absence of information on the criteria determining the selection of the “unseen data” used to evaluate the “model’s predictive performance”. 	<ul style="list-style-type: none"> ● The apparent ‘generalisability’ of the ‘predictive’ patterns poses a modest risk of misinterpretation as mechanistic (and hence ‘causal’) were customers to mistake algorithmic/probabilistic ‘prediction’ (a type of estimation, similar in inferential value to an ‘association’) for causally definitive insight. 	<p><i>“The performance of the model across unseen datasets is consistent across folds, suggesting that the patterns identified are generalizable to these datasets.”</i></p>

Bootstrapping

Bootstrapping

“Across thousands of resampled datasets, the key estimate remains tightly clustered, suggesting the conclusion is robust and not driven by a small number of observations.”

- The thematic labelling of this technique is appropriate.
- Citing “thousands of resampled datasets” (not unusual for this technique) is superfluous and misrepresents the assiduousness of its application.
- Alternative sources of non-robust clustering (beyond “a small number of observations”) warrant including in such summaries of this technique where relevant.
- The fact that bootstrapping offers some reassurance that “key estimate(s)” of smaller sample sizes have been appropriately quantified (always) poses a modest risk that such estimates are mistaken as representative of the population from which the “small number of observations” were drawn – a sample size that nonetheless poses a substantive risk of being unrepresentative.

“Bootstrap resampling yielded a narrow sampling distribution for the estimate – indicating that, despite the small number of observations, the key estimate generated from the sample had substantial statistical precision.”

5.1. Do These “Typical Analytical Insights” Methodologically Misrepresent the Techniques Involved?

From Table 2 it is clear that almost all of the “typical analytical insights” – as originally drafted by Duffield (2026) – misrepresent the methodological intricacies and conditionalities of each of the techniques concerned, albeit to a varying degree. For the most part the level of misrepresentation may appear relatively trivial, or potentially justifiable on the basis that the less cautious wording used was deemed necessary to simplify and communicate these “insights” to non-specialist audiences (including intelligence analysts or customers yet to be trained in each of the 11 statistical analytical techniques). Yet all too often the drafted “insight” omits or unhelpfully obfuscates inferentially crucial methodological details, such as:

- the intended ‘analytical objective’ of the “typical” example/application used as a basis for drafting each technique’s “insight” (i.e. whether the objective be for descriptive, ‘predictive estimation’, optimisation and/or causal inference purposes) – an issue affecting *all* of the examples/applications and “insights” drafted for these 11 techniques; and a particular issue for those techniques that can be modelled in different ways to inform more than one analytical objective;
- which covariates were adjusted for/conditioned upon – as for: the “other variables” that were “controll[ed] for” in the “insight” drafted for Linear Regression; and variables relevant to “baseline capability and environment” that were “accounted for” in the “insight” drafted for ANCOVA;
- what *other* additional/subsidiary/ancillary techniques were involved to generate *additional* findings that would not have been generated by the principal technique concerned – as for: the “threshold effect” in the “insight” drafted for Logistic Regression; the “natural” separation of “distinct groupings” (and the “material” nature of these groupings’ “risk profiles”) in the “insight” drafted for K-Means Clustering; the “[statistical] significance” of the “narrow[ing]” of differences between “[different groups of] actors” in the “insight” drafted for ANCOVA; the “sharp[ness]” of the “deviation” in the “insight” drafted for ARIMA; and the “structural” nature of the “longer term... change” in the “insight” drafted for Trend Plotting;
- whether any extrapolated estimates of future phenomena/trends might be subject to the assumption of *ceteris parabus* (all other things being equal/remaining unchanged) – as would be the case for the risk-associated inference of “actor” classification inferred in the “insight” drafted for K-Nearest Neighbours; and
- the meaning of value-laden claims (and, where possible, the quantification of such claims) – as for the “bounded range” and “non-trivial[ity]” of trends in “activity levels” in the “insight” drafted for ARIMA; the “substantially... increased... confidence” and “materially lower probability” in the “insight” drafted for Bayesian Analysis; and the “sharp deterioration” in the “insight” for Monte Carlo Simulation.

Besides these omissions and obfuscations, there was also one instance where the technique had been substantively misrepresented – this being the “causal inference” thematic label applied to the Analysis of Covariance (ANCOVA). While this technique *can* support models capable of generating (probabilistic) causal inference (as well as models whose analytical objective is ‘predictive estimation’ or optimisation) so can other, *predominantly correlational/associational* techniques, such as Linear and Logistic Regression – both of which, in this instance, had been thematically labelled *methodologically* as “regression”, rather than *inferentially*, as “causal inference” and/or ‘predictive estimation’.

That said, the “causal inference” thematic label applied to ANCOVA (and only thereto), did not appear to have substantively undermined the “typical analytical insight” drafted for this technique. Indeed, its summary of the impact of “account[ing] for...baseline capability and environment[-related covariates]” on the “apparent gap between [groups of] actors” is largely coherent from a methodological and inferential point of view. The “insight” drafted for this technique nonetheless failed to mention the additional/subsidiary/ancillary technique(s) that would have been required to

determine whether the “narrow[ing of the]... gap” was “[statistically] significant” – something that ANCOVA alone does not routinely or directly assess.

Taken in the round, these misrepresentations fall substantively short of the more comprehensive, carefully worded and detailed descriptions that might otherwise have ensured that each of the drafted “insights” offered unambiguous yet accessible/non-technical summaries of what each of the 11 techniques were able to provide.

5.2. Do these “typical analytical insights” lead to the inferential misinterpretation of the techniques involved?

As a result of the omissions, obfuscations and misrepresentations summarised in Section 5.1 and Table 2 (above), many of the “typical analytical insights” (as originally drafted) present multiple opportunities for (potential) inferential misinterpretation. These include those: that reflect the paucity of methodological information available to fully understand precisely how each of the techniques (and any additional/subsidiary/ancillary techniques) had been applied; and where the inferences offered/implied by the “insights” (as drafted) go some way *beyond* what the techniques concerned are able to support:

- Linear Regression – This analytical technique *can* be used to support causal inference based on bivariate associations evident within observational datasets (in this instance that: “X is a... driver [for Y]”) but *only* under very specific conditions (these being: temporal-consistency; dataset representativeness; and the availability, and selection, of an appropriate covariate adjustment set; see ANCOVA, below for more details). Even when these conditions hold, this technique is *not* able to determine *definitively* whether any such association is “meaningful”, since the specified cause/“driver” (in this instance: “X”) may only be incidentally/indirectly *associated* with a *direct* cause of Y (since such an association can also occur through chance, unacknowledged collider bias, or as a result of unadjusted/unmeasured/residual confounding). The “insight” as drafted therefore warrants revision to: temper these explicit inferential claims (and their narrow, and potentially unfounded/unsound, interpretations); or better detail the methodological dependencies upon which these rely (including each of the specific conditions – temporal-consistency, dataset representativeness, and appropriate adjustment set composition – necessary for generating probabilistic causal inference from observational datasets; *and* any *additional* techniques applied to assess the “meaningful”-ness of the causal link inferred between “X” and “Y”).
- Logistic Regression – This analytical technique *can* be used to determine whether the ‘odds’ of a binary outcome (in this instance: the “probability of event A” *occurring vs. not occurring*) is higher/lower, stronger/weaker or more/less precise in the presence/absence of other “factors”. However, this technique neither provides nor supports inferential assessment as to whether a “threshold effect” is/is not present, or what “factors” might be associated with (or responsible for) such an effect’s ‘inflection point’, *unless* the model(s) used is(are) designed so as to support such assessments. It is also at risk of inferential misinterpretation were “factors X and Y” to have occurred/crystalised *after* “event A” happened (i.e. as potentially direct/indirect *consequences* of “A”); *before* “event A” (i.e. as potentially direct/indirect *causes* of “A”); or, indeed, were these “factors” to be ‘mathematically coupled’ to, or indivisible components/features of, “event A”. The apparent relationship between (the presence/absence of) “X” and “Y” and the odds of “event A” will likewise only be consistent with a causal relationship under very specific conditions (these being: temporal-consistency; dataset representativeness; and the availability, and selection, of an appropriate covariate adjustment set; see ANCOVA, below for more details). The “insight” as drafted therefore warrants revision to: temper these inferential claims (and their narrow, and potentially unfounded/unsound, interpretations); or better detail the methodological dependencies upon which these rely (including any *additional* modelling/analytical techniques applied to assess the: nature of the relationship between “X” and “Y” and “event A”; and presence of a potential “threshold effect”).

- K- Nearest Neighbours – This analytical technique *can* be used to classify/group entities, events, processes or characteristics thereof with regard to similarities in related features (in this instance: “recent behaviour”). However, it does not necessarily provide or support inferential assessment as to *whether* any group member so classified necessarily shares the *subsequent* risk profile of all/most of the other group members (in this instance: “a small subset of previously observed cases”), unless: such inference involved risk-relevant outcomes (in this instance: “outcome B” or causes/determinants thereof) that were components of, mathematically coupled to, or (manifestly/latently, and directly/indirectly) causally associated with, the variables used for classification/grouping; *and* all necessary and sufficient factors/circumstances required for the “historical... outcome[s]” concerned to remain in place. The “insight” as drafted therefore warrants revision to: temper these inferential claims (and their potentially unfounded/unsound interpretations); or better detail the methodological dependencies upon which these rely (including the inclusion of risk-relevant/related components/correlates/latent features within the classification models; and the temporal stability of any factors/circumstances on which “outcome B” depends).
- K-Means Clustering – This analytical technique *can* be used to classify/group entities, events, processes or characteristics thereof with regard to similarities in related features (in this instance: “recent behaviour”). However, it does not necessarily provide or support inferential assessment as to: *why* the groupings are evident or *how* they arise (in this instance: whether “naturally” vs. artefactually); or the *nature/extent* of any group-specific meta-features (in this instance: “internally consistent patterns”) or group-related differences in risk profiles (in this instance: whether “material” vs. intangible/insubstantial), unless: such inference involved risk profile-relevant/related features that were components of, mathematically coupled to, or (manifestly/latently, and directly/indirectly) causally associated with, the variables used for classification/grouping; and all necessary and sufficient factors/circumstances required for the “risk profiles” concerned to remain in place. The “insight” as drafted warrants revision to temper these inferential claims (and their potentially unfounded/unsound interpretations); or better detail the methodological dependencies upon which these rely (including the inclusion of: what techniques/measures were used to determine whether the groupings “separate... naturally”, and the “internally consistent patterns” and “materially different risk profiles” described; and the inclusion of risk-relevant/related components/correlates/latent features within the classification models).
- ANOVA (Analysis of Covariance) – As its “thematic classification” suggests, this analytical technique *can* be used to inform (probabilistic) causal inference (as can other, predominantly associational/correlational analytical techniques, such as Linear and Logistic Regression; see above) but only – as mentioned earlier – under very specific conditions. These are: temporal-consistency (such that: the specified cause precedes its specified consequence; and any covariates controlled/adjusted for/conditioned upon precede *both* the specified cause and consequence); dataset representativeness (to mitigate the risk of a form of collider bias, known as ‘endogenous selection bias’; Elwert and Winship, 2014); and the availability, and selection, of an appropriate, and appropriately diverse, covariate adjustment set (so as to optimally mitigate the risk of measured/measurable confounding, while avoiding any additional risk of collider bias resulting from inappropriate adjustment for variables that occur *after* the specified cause and/or consequence; Ellison, 2023). In the “insight” originally drafted for this technique, the risk of inferential misinterpretation is modest, despite the technique being thematically misclassified as (predominantly/solely) “causal inference”. Nonetheless, the allusion to a narrowing of an “apparent gap” between (presumably) different *groups* of “actors” after “accounting for... baseline capability and environment” does imply that the latter were considered potential sources of confounding that, once “accounted for” (presumably through inclusion of relevant variables in the model’s covariate adjustment set), revealed a much narrower gap between the actors – suggesting that (groups of) different “actors” were more similar in terms of their (more

recent) “behaviour” than was “apparent” prior to adjustment. This constitutes causal inference in the sense that the (confounding-adjusted) differences between “actors” was found to be small, and therefore predominantly caused by preceding differences in “baseline capability and environment” (i.e. factors acting as bona fide ‘confounders’) rather than contemporaneous divergence in behaviour. Meanwhile, the use of the phrase “narrows significantly” implies that an additional statistical technique will have been used to assess the statistical significance of the reduction observed following adjustment – an assessment that ANCOVA does not ordinarily or *directly* provide (unless specifically modelled and configured to support this). For these reasons, the “insight” as drafted warrants revision to better qualify these inferential claims (not least with more detail concerning the nature of the “apparent gap between actors”) so as to facilitate greater understanding of what role “divergent behaviour” might play, whether as a specified consequence of “actor” identity, or as a mediator on the causal path between “actor” identity and a separately specified consequence; and better detail the methodological dependencies upon which these rely (including the inclusion of: what techniques/measures were used to assess how the “gap between actors narrow[ed] significantly” following adjustment for “baseline capability and environment”).

- Trend Plotting – While often mistaken for a purely descriptive, univariable technique, Trend Plotting is essentially a bivariable analysis that helps reveal any sequential variation in a specified target/outcome variable in relation to the changing value of a second ordinal or continuous variable (such as a ranked feature or, more commonly, time). It is capable of elucidating and mapping past/current patterns in this variation and thereby offering a retrospective/historical basis upon which past variation might be better characterised (and – potentially – the mechanisms responsible better understood, and applied to generate ‘foresight’ of possible future trends and the ‘predictive estimation’ of dataset features through extrapolation. In the drafted “insight” provided, the inferential claims appear modest since they simply describe the presence of an historical “sustained... longer term... upward movement” that is otherwise, to some degree, obscured by “short-term volatility” – claims that are based solely on extant empirical information. However, where this “insight” does stretch the inferential capabilities of the technique concerned is in suggesting that such (historical) patterns can “show... [evidence of] structural change”. While the specific example and context to which this “insight” refers might potentially clarify/justify such a claim, the omission of this information from the text of the “insight” implies a level of mechanistic/structural understanding that goes beyond what the technique itself can support. The “insight” as drafted therefore warrants revision to temper these inferential claims (and their potentially unfounded/unsound interpretations); or better detail the contextual/situational and methodological dependencies upon which these rely (including the inclusion of: why the “sustained upward movement” is felt to reflect a longer term “structural change” as opposed, for example, to a regular and sustained series of inputs [of some form or another] that might plausibly have elicited such a trend; and so on).
- ARIMA (Auto-Regressive Integrated Moving Average) – In common with Trend Plotting (see above), this is a bivariable technique that can help reveal any sequential variation in a specified target/outcome variable in relation to the changing value of a second ordinal or continuous variable (such as a ranked feature or, more commonly, time). It is similarly capable of elucidating and mapping past/current patterns in such variation and thereby offering a retrospective/historical basis upon which past variation might be better characterised (and – potentially – the mechanisms responsible, and any possible future trends, speculatively inferred, understood or estimated, respectively). Such future trends are estimated through mathematical/statistical extrapolation (commonly, if unhelpfully, called ‘prediction’), albeit under the assumption of *ceteris parabus* (all other things being equal/remaining unchanged). Indeed, the drafted “insight” provided for this technique begins with a disclaimer to this effect (“assuming current dynamics persist”) before characterising the estimated future trend in

“activity levels... over the next three periods [of time]”. In this regard the inference implied appears both clear and defensible, though the absence of any qualification as to what is meant by: the future trend “remain[ing] within a bounded range” (the meaning of which is *entirely* predicated on the size/scale of the “range” concerned); or a “non-trivial risk” and a “sharp deviation” – substantively detracts from the import of any inference implied. The “insight” as drafted therefore warrants revision to better clarify and qualify these inferential claims (and their potentially unfounded/unsound interpretations); by offering more detail on any quantifiable measures of the “bounded range”, “trivial risk” and “sharp deviation” to which it refers.

- Bayesian Modelling – This (broad group) of techniques, in which ‘prior’ hypotheses concerning the distributional and correlational properties of datasets are tested against new information/data (as this becomes available), offer an attractive alternative to more commonplace ‘frequentist’ techniques – particularly for those analysts who are keen to test their prior understanding (whether speculative or confident) of the contexts and mechanisms in which, and through which, their analytical objectives/problem sets arise. In the “insight” drafted for this technique, the inference is that “new reporting” has confirmed the hypothesised prior(s) and thereby increased the “[analytical] confidence” therein (while “lower[ing the] probability [or likelihood that]... alternative explanations” [are sound]). Whether the technique (as applied in this instance) supports such inference will depend in no small part on the basis upon which the ‘prior’ was derived, and whether the “new reporting” constituted a contribution that was: independent of whatever empirical/theoretical/speculative evidence led to the analyst’s ‘prior’; and capable of challenging/falsifying/qualifying their initial hypothesis. Provided both of these conditions hold, the “insight” as drafted should not lead to inferential misinterpretation – although the use of the term “confidence” to reflect the finding that the ‘prior’/hypothesis was consistent with “new reporting” is unfortunate given “confidence” is more commonly used (in statistical parlance) with reference to precision, and its use here might therefore prompt misinterpretation. Likewise, the use of “materially” to describe (and thereby substantiate) the “lower probability [of]... alternative explanations” would benefit from more detailed (and, where possible/appropriate, quantitative) qualification, alongside further detail on any additional/subsidiary/ancillary techniques used to assess this. The “insight” as drafted therefore warrants revision to better clarify and qualify these inferential claims (and their potentially unfounded/unsound interpretations); by offering more detail on: how the ‘prior’ hypothesis was developed, and in what sense the “new reporting” might constitute an appropriately rigorous/robust test thereof; and what techniques/measures were used to determine “confidence in hypothesis A”, and the “material[ity]” of the “lower probability” observed for the “alternative explanations”.
- Monte-Carlo Simulation – Though lacking in methodological detail (and notwithstanding the unnecessary reference to “thousands of simulations”, given these are germane to this technique), the drafted “insight” in this instance explicitly invites inference regarding the four “plausible” scenarios supported by the distributional and correlational properties of the (resampled) dataset available. Assuming, once more, that the resampling and algorithmic procedures were appropriately applied, such inference is nonetheless predicated upon: precisely how (and with what criteria) “emerg[ing]... scenarios” were considered “plausible”; why only one of these scenarios was deemed necessary/relevant to describe in any (additional) detail; and how both a “sharp” change and a “deterioration” (as in “sharp deterioration”) were defined, identified and classified as such. The “insight” as drafted therefore warrants revision to better clarify and qualify these inferential claims (and their potentially unfounded/unsound interpretations); by offering more detail on: the methodological procedures employed; all four of the “plausible” scenarios these procedures identified; how the “plausib[ility]” of “emerg[ing]... scenarios” was determined (particularly if any statistical tests/parameters informed their determination); and how both “sharp” and “deterioration” (and the possible converse, be these: absent or modest; and stability or improvement, respectively) were defined and determined (including, as before,

detail of any additional/subsidiary/ancillary techniques statistical tests and parameters used to judge such assessments).

- Cross-Validation – This technique’s “insight” – assuming, once more, that the choices/decisions made were appropriate to the dataset, context and analytical objective/problem set concerned – explicitly infers that the original model can be considered appropriate to the dataset examined simply because its predictive performance “remains stable” when tested in subsidiary (hence “unseen”) datasets *other than* that/those used to ‘train’ the initial model(s)/algorithm(s). While such findings *are* consistent (and therefore *reasonable*) evidence that the original model was “generalisable” and unlikely to have been “overfitted” to the original ‘training’ dataset, such inference substantively depends upon: the complexity of the dataset; and the inherent variability of the distributional and correlation properties of its constituent variables – and whether *both* the dataset- *and* data-generating mechanisms are inherently stable (i.e. structurally and statistically conserved, regardless of context – as can be the case with artefacts and contexts governed by tight ‘rules’ or design-related constraints). In such instances, this technique may prove less insightful than the inference implied might suggest – *unless*, that is, prior knowledge/understanding of either/both of these mechanisms was absent, incomplete, imprecise, inaccurate, unreliable or invalid (which may often be the case in intelligence analysis). The “insight” as drafted therefore warrants revision to better clarify and qualify these inferential claims (and their potentially unfounded/unsound interpretations) by offering (or referring back to) more of the detail on: the dataset(s), dataset variables and context(s) examined; and adding further evidence regarding the “patterns... identified” so as to support further inferential interpretation/speculation thereon.
- Bootstrapping – Setting aside the unnecessary reference to “thousands of resampled datasets” – which (like the “thousands of simulations” within the “insight” drafted for Monte Carlo Simulation; see above) is germane to this technique – the “insight” drafted suggests it did not achieve the principal benefit for which Bootstrapping is most commonly applied – this being to generate *more* precise (i.e. “[more] tightly clustered”) “key estimate[s]” than would otherwise be possible in datasets containing a relatively “small number of observations” (given the higher risk of chance sampling/measurement imprecision therein). Since, in this instance, the “key estimate *remains* tightly clustered” (*italicised* emphasis added), the self-evident inference on this occasion is that the *original* level of “cluster[ing]” (or precision) observed across the whole of the dataset – given this had a relatively “small number of observations” – does *not* improve *despite* the application of bootstrapping. Alternatively, one might assume that the analytical objective behind the use of this technique (on this occasion) was to challenge whether a more “tightly clustered... key estimate” than was initially considered possible/likely might have been a chance phenomenon given the “small number of observations” from which this “key estimate” was derived. Yet bootstrapping is unlikely to offer a robust assessment of the potential for ‘Type 1’ errors of this sort, since “tightly clustered... key estimate[s]” generated initially using the entire dataset are unlikely to prove *less* “tightly clustered” (i.e. less *precise*) following bootstrapping simply because, by definition, the scope for *imprecision* is clearly *already* limited. The “insight” as drafted therefore warrants revision to better clarify and qualify these inferential claims (and their potentially unfounded/unsound interpretations) by including additional information relevant to the analytical objective the technique was deployed to address.

6. Are Duffield’s 11 Techniques Suitable for Intelligence Analysts with “a Relatively Low Mathematical Baseline”?

It is clear from the preceding Sections of this article that substantial knowledge, care and expertise is required to competently apply (see 1 through 4, above), and succinctly summarise (see 5, above) the 11 “statistical analytical techniques” that Duffield (2026) proposed would “form a strong baseline for... intelligence analysis”. While he himself acknowledged that many intelligence analysts have “a relatively low mathematical baseline”, the scale of the gap between their current statistical

competencies and those required to confidently and competently apply each of his 11 proposed technique warrants closer examination – and not least were this to be a relatively narrow gap, and thereby easily addressed through training and/or workplace-based experience and supervision.

Unfortunately, a definitive answer to this question is likely to be elusive, and not simply because different intelligence services/agencies apply different (and often deliberately opaque) criteria when recruiting and developing analysts; and these criteria depend upon the nature of the specialist and generalist qualifications, aptitudes and experience that different intelligence (collection) disciplines and (analytical) specialisms feel the need to apply (MOD, 2023). Some country's specialist agencies – such as Israel's 'Unit 8200' (which, it is claimed, has been responsible for the rapid proliferation of data science and AI-enabled collection and targeting algorithms during the ongoing conflict in the Middle East; see Yuval, 2024) – go to great lengths to identify recruits with the mathematical, computational and contextual skills, aptitudes and interests deemed necessary for algorithmic innovation (Corera, 2016; Thomas and Kruppa, 2024). However, most entrants into generic intelligence analysis roles do not undergo such targeted selection/recruitment. It is therefore against this lower bar that it is necessary to assess the scale of the gap were these personnel to apply the “core set of [11] statistical analytical techniques” proposed (see Table 1, above).

For want of more accessible example, the current baseline entrance requirements for applicants to intelligence trades of within the UK's armed forces (see: RN, 2026a; British Army, 2026; and RAF, 2026) are based either on successful completion of a dedicated aptitude test (such as the Royal Navy's Defence Aptitude Assessment; RN, 2026b) or a Grade 4 (i.e. a 'standard pass') or above in GCSE Mathematics. In contrast, entrants into the UK's armed forces as intelligence officers, or the UK's civil service intelligence profession, are commonly required to have higher grades in mathematics (either at GCSE or A-Level); while some specialist intelligence disciplines and data-intensive civil service/intelligence agency careers often require entrants to have graduate-level qualifications in mathematics (or substantive evidence/experience of comparable data handling and analysis skills). These higher qualifications aside, it nonetheless seems likely that a good many of those entering intelligence analysis roles in the UK – and therefore those to whom Duffield's (2026) argument would apply – will only have achieved a 'standard pass' in GCSE Mathematics.

Assuming this constitutes the “relatively low mathematical baseline” to which Duffield (2026) refers, the content of the UK's GCSE Mathematics course should provide a clear indication of the skillsets that those who gain a 'standard pass' should possess. This is somewhat complicated by the range of 'GCSE Mathematics' courses available. These include: two labelled “GCSE Mathematics” – one at a “Foundation Tier” (offering Grades from 1-5), the other at a “Higher Tier” (offering Grades from 4-9); and four offering a more focussed, or higher Tier, mathematical qualification – these being “GCSE Statistics”, “Level 2 Certificate in Further Mathematics”, “Level 2 EMC (Extended Maths Certificate)”, and “Level 3 FSMQ (Free Standing Mathematics Qualification in Additional Mathematics)”. However, since none of these four (more focussed or higher Tier) mathematics qualifications are required for entry into the intelligence branch of the UK's armed forces, the “baseline” Duffield (2026) describes will be equivalent to a Grade 4 'standard pass' (or above) in either the “Foundation Tier” or “Higher Tier” of GCSE Mathematics.

Both Tiers cover 6 broad topics (core numeracy; basic algebra; ratios; geometry; probability; and statistics) – albeit at a shallower (Foundation Tier) or deeper (Higher Tier) level. The last two of these (probability and statistics) are the most relevant for assessing Duffield's (2026) premise, and the specific content of each of these (P1-P9; and S1-S6) has been summarised in Table 3a,b, below. Clearly, neither of these topics are taught at a level that would equip those who achieve a 'standard pass' to attain the more technical, analytical and inferential, competencies summarised in Tables A1.1-A4.2 (see Appendices) without recourse to an extended period of training and/or specialist workplace-based supervision and support. The substantive and opportunity costs associated with such training, supervision and support (and the continuing professional development required to maintain/refresh/update such expertise) are far from insubstantial; and even were the time and resource required to be available, there is clearly a risk that a good number of entrants who currently

qualify as competent and proficient intelligence analysts might fail to achieve the mathematical and statistical skills required to take on these *additional* responsibilities.

Table 3a. Summarised content on “Probability” (hence P1-P9) for the 2026 GCSE Mathematics curriculum, relevant to the Foundation and Higher Tier where appropriate (OxfordAQA, 2026).

Section	Basic foundation content	Additional foundation content	Higher content only
P1	Record, describe and analyse the frequency of outcomes of probability experiments using tables and frequency trees.	[Blank]	[Blank]
P2	Apply ideas of randomness, fairness, and equally likely events to calculate expected outcomes of multiple future experiments.	[Blank]	[Blank]
P3	Relate relative expected frequencies to theoretical probability, using appropriate language and the 0 to 1 probability scale.	[Blank]	[Blank]
P4	Apply the property that the probabilities of an exhaustive set of outcomes sum to 1.	[Blank]	[Blank]
P5	[Blank]	Understand that empirical unbiased samples tend towards theoretical probability distributions, with increasing sample size.	[Blank]
P6	Enumerate sets and combinations of sets systematically, using tables, grids, and Venn diagrams.	[As for Basic, but also] including using tree diagrams.	[Blank]
P7	Construct theoretical possibility spaces for single and combined experiments with equally likely outcomes; and use these to calculate theoretical probabilities.	[Blank]	[Blank]
P8	Calculate the probability of independent and dependent combined events, including using tree diagrams and other representations; and know the underlying assumptions.	[Blank]	[Blank]
P9	[Blank]	[Blank]	Calculate and interpret conditional probabilities through representation using expected frequencies

			with two-way tables, tree diagrams and Venn diagrams.
--	--	--	---

^aP1 Notes: probabilities should be written as fractions, decimals or percentages. ^bP8 Notes: including knowing when to add and when to multiply two or more probabilities.

Table 3b. Summarised content on “Statistics” (hence S1-S6) for the 2026 GCSE Mathematics curriculum, relevant to the Foundation and Higher Tier where appropriate (OxfordAQA, 2026).

Section	Basic foundation content	Additional foundation content	Higher content only
S1	[Blank]	Infer properties of populations or distributions from a sample, whilst knowing the limitations of sampling.	[Blank]
S2 ^c	Interpret and construct tables, charts and diagrams, including: frequency tables, bar charts, pie charts and pictograms for categorical data; vertical line charts for ungrouped discrete numerical data; and know their appropriate use.	[As for Basic, but also] including tables and line graphs for time series data.	[Blank]
S3	[Blank]	[Blank]	Construct and interpret diagrams for grouped discrete data and continuous data – i.e. histograms with equal and unequal class intervals, and cumulative frequency graphs; and know their appropriate use.
S4 ^d	Interpret, analyse and compare the distributions of data sets from univariate empirical distributions through: <ul style="list-style-type: none"> ● appropriate graphical representation involving discrete, continuous and grouped data; and ● appropriate measures of central tendency (median, mean, mode and modal class) and spread (range, including consideration of outliers). 	[Blank]	[As for Basic, but also:] <ul style="list-style-type: none"> ● including box plots; and ● including quartiles and inter-quartile range.
S5	Apply statistics to describe a population.	[Blank]	[Blank]
S6 ^e	<ul style="list-style-type: none"> ● Use and interpret scatter graphs of bivariate data; and ● Recognise correlation. 	[As for Basic, but also:] <ul style="list-style-type: none"> ● know that it does not indicate causation ● draw estimated lines of best fit 	[Blank]

		<ul style="list-style-type: none"> ● make predictions ● interpolate and extrapolate apparent trends whilst knowing the dangers of so doing 	
--	--	--	--

²S2 Notes: including choosing suitable statistical diagrams. ³S4 Notes: students should know and understand the terms: primary data, secondary data, discrete data and continuous data. ⁴S6 Notes: students should know and understand the terms: positive correlation, negative correlation, no correlation, weak correlation and strong correlation.

Conclusion

Jack Duffield's (2026) proposition that (all) intelligence analysts should become competent in "a core set of statistical analytical techniques" is flawed and potentially dangerous:

- first, because it is far from clear that these techniques are required to address (or would succeed in addressing) the "total information overload" he believes they face due to the accelerating "volume, variety and velocity" of so-called 'Big Data' (Ellison, 2026); and
- second, because the scale of the training, supervision and support required to upskill intelligence analysts in the 11 "statistical analytical techniques" appears impracticable and implausible without substantial additional time and resource, and a concomitant recalibration of current intelligence analysis doctrine and practice.

Were no additional resource and recalibration to be available or implemented, requiring (or even encouraging) intelligence analysts to take on these (more advanced) statistical responsibilities risks vitiating the intelligence analyses they generate with multiple – potentially damaging, yet avoidable and arguably unnecessary – analytical and inferential mistakes.

There may nonetheless be an argument for establishing a dedicated specialist sub-discipline with the expertise required to exploit recent technological advances in data analytics and generate robust insights from the statistical analysis of large, novel quantitative datasets (including 'Big Data'). But non-specialist intelligence analysts should be discouraged from attempting to apply specialist analytical techniques – including advanced statistical analysis – without the training, expertise and competencies required to avoid unnecessary (and potentially dangerous) mistakes.

Endnotes

¹ Statistical analysis is, at heart, simply an attempt to extract useful insight from the central tendencies, patterns and trends present within (medium- to large-scale) quantitative datasets, while simultaneously mitigating any non-systematic error (i.e. imprecision or 'noise') resulting from 'measurement error', and addressing the most important sources of systematic error (i.e. bias).⁴

² That is, to support the estimation/optimisation of hypothesised or (as yet) unmeasured dataset features (whether past, present or future) using univariable or multivariable interpolation, extrapolation and latent class/variable techniques.

³ Such as how comprehensively a 'predictive estimation' algorithm captures the distributional statistical information available, or how sensitive the analyses might be to modest levels of non-systematic and systematic error in the measurement of key variables (and particularly so for the target/outcome variable; see Section 4).

⁴ The reliance of statistical analysis on substantive *samples* of data (comprising *multiple* measurements/observations or measurements/observations from *multiple* cases) is a consequence of the way this works by: exploiting and interrogating the distributional and

correlational properties of *multiple* measurements/observations; in order to identify central tendencies, patterns and trends from amongst the ‘noise’ of ‘measurement error’.

Appendices

Appendix A: A non-technical critique of Duffield’s (2026) proposition

The following critique has been accepted for publication in *The RUSI Journal* ([DOI](#)):

Statistics in Intelligence Analysis: “a little learning is a dang’rous thing”

A little learning is a dang’rous thing; Drink deep, or taste not the Pierian spring; There shallow draughts intoxicate the brain; And drinking largely sobers us again (Alexander Pope 1711; 216).¹

Jack Duffield² argues that intelligence analysts face “total information overload” due to the accelerating “volume, variety and velocity” of so-called ‘Big Data’.³ Yet *excessive* amounts of quantitative (and enumerable) data were being generated long before advances in the digitisation of information and communication picked up pace in the early 1990s.³ Decades earlier, as Duffield himself concedes, MASINT had emerged as a specialist intelligence subdiscipline precisely to support “the analysis of [the larger and more complex] data[sets] obtained from sensing instruments”⁴ – datasets that could not be adequately exploited by the analytical techniques then in use by other subdisciplines. It is nonetheless questionable whether *all* intelligence analysts now need additional *data* analytic skills to deal with the risk of “overload” from ‘Big Data’ – even if only in the form of a “basic toolkit of... Statistical Analytical Techniques”.² This is because:

- ‘Big Data’ seems unlikely to be the most pressing cause of analyst “overload” – given analytic tradecraft routinely necessitates extended periods of intensive exposure to computer-based digital workspaces, characterised by a relentless stream of multi-sensory/multi-tasking demands, each capable of overwhelming their cognitive processing capacity;⁵
- only analysts working within specialist “collection disciplines” or “analytical specialisms”⁴ that *require* the (statistical) analysis of quantitative datasets will necessarily benefit from enhanced *statistical* competencies – and those that do may benefit little from the “basic toolkit”² proposed;
- burdening *all* analysts with *additional* responsibilities to develop and apply *additional* statistical skills may exacerbate any “overload” they experience – unless these responsibilities were offset by a commensurate uplift in analytical capacity/efficiency, or a reduction in analytical output/productivity; and
- generating robust, statistically derived insights from quantitative data requires substantive understanding of the many sources of non-systematic and systematic error (bias) – both analytical *and* inferential – that arise at each and every step in the application of even the most “basic statistical analytical techniques”.²

The last of these appears most troublesome. This is because strengthening “basic” statistical knowledge – on its own – risks enabling (if not encouraging) intelligence analysts to apply such knowledge beyond the bounds of its limited utility;⁶ thereby vitiating intelligence products with unreliable and invalid insights imbued with the misplaced authority commonly afforded quantitative findings.⁷ Duffield illustrates this by oversimplifying (and occasionally misrepresenting or misinterpreting) *all* of the “typical insights” provided by the 11 “core statistical analytical techniques” summarised in Table 1 of his article.^{2,8} While this may partly reflect the brevity imposed by the space constraints of an academic article, each of his “typical insights” *also* reflect many of the commonest analytical/inferential mistakes made by quantitative researchers – including statistically literate analysts – at some point in their careers (it being all too easy to overlook sources of error and bias if you are unfamiliar with how these arise and should be addressed).⁹ Duffield and I (being one such analyst) are in good company in this regard – though neither of us will find much comfort in that, given the centrality of rigorous methodology to professional intelligence practice.

Analysts keen to learn more will need to look beyond the introductory statistical textbooks that Duffield cites, and seek the expertise of statisticians familiar with mitigating the risks of non-systematic and systematic error (bias) involved at each of the preparatory, analytical and inferential

steps involved in: data sampling, acquisition, measurement and processing; the selection, design and application of the analytical techniques and modelling used; and the interpretation of, and inferences drawn from, their analytical outputs. In the meantime, the intelligence profession might best be served by:

1. Adopting cognitive hygiene practices as part of its standard operating procedures, to reduce the “overload” attributable to ‘technostress’ (in which ‘Big Data’ *may* play a part);^{5,10}
2. Establishing a dedicated specialist sub-discipline with the expertise required to generate robust insights from the statistical analysis of large quantitative datasets (and ‘Big Data’);
3. Ensuring that all analysts understand the inferential pitfalls that often accompany error and bias in statistical analysis (whenever these are not avoided, attenuated, mitigated, or acknowledged and accommodated);⁹ and
4. Discouraging analysts from attempting specialist analytical techniques – such as statistical analysis – without the training, expertise and competencies required to avoid unnecessary mistakes.

George Ellison is Professor of Data Science in the Centre for Intelligence Studies at the University of Lancashire, gthellison@lancashire.ac.uk

¹Pope A. An essay on criticism (W Lewis: London, 1711). In: *Literature in Context: An Open Anthology*. Archived [here](#) on 18APR24

²Duffield J. Statistical analytical techniques for intelligence analysis. *RUSI Journal* 2026; 171: 1-12. DOI: 10.1080/03071847.2026.2646065

³ Beer D. How should we do the history of Big Data? *Big Data and Society* 2016; 30: 2053951716646135. DOI: 10.1177/2053951716646135

⁴ Ministry of Defence. *Intelligence, Counter-intelligence and Security Support to Joint Operations*. Development, Concepts and Doctrine Centre, Ministry of Defence, UK; 2023: 181pp. Archived [here](#) on 20NOV23

⁵ Correia de Barros E. Understanding the influence of digital technology on human cognitive functions: a narrative review. *IBRO Neuroscience Reports* 2024; 17: 415-22. DOI: 10.1016/j.ibneur.2024.11.006

⁶ This is evident in the four competency levels (“0: Awareness”; “1: Working”; “2: Practitioner”; and “3: Expert”) applied to “Analysis, Tradecraft & Assessment” – the third of five “technical skills” covered by the “Professional Development Framework” developed by the UK’s Professional Head of Intelligence Assessment (PHIA) – Archived [here](#) on 24JAN25. Only at “Practitioner” and “Expert” levels is there any expectation of practical statistical competence: “At Practitioner, you effectively apply probabilistic reasoning and logic to your assessments and use structured analytical techniques *and, where appropriate, quantitative methods* to produce your assessments.”; and “At Expert, you proactively and constructively challenge conventional thinking and assessment. You are *comfortable working with complex data*. You are a thought-leader for the community, scanning for *new methods of conducting analysis, tradecraft and assessment*. You *liaise with experts on intelligence analysis techniques* and promote the use of novel approaches in the community.” (*emphases added in italics*).

⁷ Wood, M. Bridging the relevance gap in political science. *Politics* 2014; 34: 275-86. DOI: 10.1111/1467-9256.12028

⁸ Due to space constraints, the oversimplification (and occasional misrepresentation or misinterpretation) of the 11 “core statistical analytical techniques” included in the “typical analytical insights” summarised in Table 1 of Duffield’s article² has been explored in greater detail elsewhere (see main manuscript here; and, in particular, Table 2 above).

⁹ Perhaps the most commonplace and important of these are: (i) mistaking correlation and association as definitive evidence of causation (see [here](#)); and (ii) misinterpreting the ‘predictive estimation’ of unmeasured *future* values using extrapolation as equivalent to causality-informed, whole system foresight (see [here](#)). See also: Good PI, Hardin JW, Hardin JW. *Common Errors in Statistics (and How to Avoid Them)*. Wiley, Hoboken, NJ; 2012: 336pp. ISBN: 0-471-46068-0; and Bracken MB. *Bias! How Systemic Error Threatens Biomedical Research*. Cambridge University Press, Cambridge, UK; 2026: 224pp. ISBN: 9781009682749

¹⁰ Rahmi KH, Fahrudin A, Supriyadi T, Herlina E, Rosilawati R, Ningrum SR. Technostress and cognitive fatigue: reducing digital strain for improved employee well-being – a literature review. *Multidisciplinary Reviews* 2025; 8: e2025380. DOI: 10.31893/multirev.2025380

Appendix B: Tables

Tables A1.1 through A4.2 in support of Sections 1 through 4 of the main manuscript:

Table A1.1. A non-exhaustive list of the conscious and deliberate, generic (technique-agnostic) choices and decisions analysts must be competent to make when using any of the 11 techniques (and *most* other quantitative analytical/statistical techniques); and the potential analytical and inferential consequences were any of these choices/decisions to be misjudged, ill-considered or made by default.

Analyst-determined choices and decisions	Consequences of misjudged choices and decisions
Selection and definition of the analytical objective	Misalignment between method and purpose; and irrelevant or misleading conclusions
Selection and definition of critical variables and features	Omitted-variable bias; spurious associations; distorted relationships; and invalid inference
How to handle/address missing data	Biased estimates; reduced power; invalid missingness assumptions; and incorrect conclusions
The minimum sample size commensurate with the intended analytical insights desired	Imprecise and biased estimates; limited capacity for trustworthy sub-group analyses; enhanced risk of Type 1 errors
Necessary data cleaning, preparation and pre-processing steps	Distorted signals; biased parameters; unstable estimates; and incorrect trends or similarity structures
Which evaluation metrics to use	Optimising for irrelevant criteria; misleading performance comparisons; and selection of inferior models
Which validation/verification approach to use	Inflated or deflated performance; information leakage; and unreliable generalisability
Which computational settings to use	Numerical instability; model non-convergence; irreproducible results; and inaccurate estimates
Which uncertainty quantification technique(s) to use	Under- or over-stated uncertainty; misleading confidence or credible/credibility intervals; and incorrect risk characterisation

Which diagnostics and robustness checks to use	Undetected misfitted model(s); unchecked and invalid assumptions; and fragile/unstable inferences
What interpretation strategy/ies to adopt	Misinterpretation of: coefficients; probabilities; forecasts, trends; and/or clusters

Table A1.2. A non-exhaustive list of ten conscious and deliberate, generic (technique-agnostic) choices and decisions analysts must be competent to make when using any of the 11 techniques (though not *all* other quantitative analytical/statistical techniques); and the potential analytical and inferential consequences were any of these choices/decisions to be misjudged, ill-considered or made by default.

Analyst-determined choices and decisions	Consequences of misjudged choices and decisions
Selecting tuning/hyperparameters	Underfitting or overfitting; poor predictive performance; distorted clusters; unstable estimates; and excessive variance or bias
Managing iterative algorithms	Failure to converge; unstable parameter estimates; sensitivity to initial conditions; and irreproducible outputs
Designing re-sampling or simulation structure(s)	Invalid interval coverage; biased uncertainty estimates; incorrect risk quantification; and unstable simulation outputs
Handling temporal or structural dependencies	Spurious autocorrelation; non-stationary residuals; poor quality forecasts; structural misspecification; and invalid validation.
Specifying probabilistic modelling components	Prior-dominated or mis-specified <i>a posteriori</i> ; incorrect uncertainty estimates; and misleading inference

Table A1.3. A non-exhaustive list of ten conscious and deliberate, particular (technique-specific) choices and decisions analysts must be competent to make when using each of the 11 techniques; and the potential analytical and inferential consequences were any of these choices/decisions to be misjudged, ill-considered or made by default.

Technique	Analyst-determined choices and decisions	Consequences of misjudged choices and decisions
Linear Regression	Functional form; handling of heteroscedasticity; identification of influential data; variance estimators; and multicollinearity treatment	Biased coefficients; invalid standard errors; poorly fitted models; and misleading inference
Logistic Regression	Link function; separation handling; correction for any class imbalance; classification threshold; and regularisation	Poor calibration; infinite or unstable estimates; skewed predicted probabilities; and misclassification
K-Nearest Neighbours	Distance metric; feature scaling; choice of K; and weighting scheme	Distorted neighbour structure; elevated bias or variance; poor discrimination; and misleading predictions
K-Means Clustering	Number of clusters; initialisation method; distance metric; algorithm variant; and stopping criteria	Wrong or unstable clusters; centroid drift; and false segmentation or grouping

ANCOVA (Analysis of Covariance)	Covariate choice; interaction specification; homogeneity-of-slopes testing; contrast coding; and error structure	Confounding; biased adjusted means; and invalid group comparisons
Trend Plotting	Smoothing method; window/span; detrending; anomaly handling; treatment of temporality and seasonality; and decomposition	Over- or under-smoothing; false patterns; and suppressed or exaggerated trends.
ARIMA (Autoregressive Integrated Moving Average)	Order selection for non-seasonal (p,d,q) and seasonal (P,D,Q) parameters; differencing choices; estimation method; structural break handling; and ARIMAX specification	Non-stationarity; over-differencing; poor forecasts; spurious autocorrelation
Bayesian Analysis	Choice of prior; hyperparameters; sampling algorithm; convergence diagnostics; and posterior predictive checks	Prior domination; non-convergence; misleading <i>a posteriori</i> inference; and incorrect uncertainty quantification
Monte Carlo Simulation	Input distributions; correlation structure and dependencies; scenario construction; and number of iterations	Unrealistic simulations; incorrect uncertainty estimates; and unstable results
Cross-Validation	Number of folds; fold construction (whether stratified, blocked or rolling); and tuning of inside/outside folds	Inflated or deflated performance; invalidation for time series; and misleading model comparisons
Bootstrapping	Bootstrap type; re-sampling unit; number of bootstrap samples; and interval type	Wrong coverage; unreliable intervals; invalid dependence structure; and incorrect uncertainty assessment

Table A2.1. A non-exhaustive list of the generic (technique-agnostic) parametric/non-parametric assumptions that need to apply when using any of the 11 techniques (and *many/most* other quantitative analytical/statistical techniques); and the potential analytical and inferential consequences were any of these assumptions not to hold.

Parametric/non-parametric assumptions	Consequences were the assumptions not to hold
Data representativeness: sample reflects the population or phenomenon being analysed	Biased estimates; invalid generalisability; and systematic distortion of inference
Measurement validity: variables accurately capture the constructs of interest	Misinterpretation; spurious associations; and misleading parameter estimates
Measurement reliability: repeated measurements and observations provide consistent values	Increased variance; attenuation of effects; loss of power; and noise-dominated inference
Correct data type usage: continuous, ordinal, categorical vs. time-indexed, where appropriate	Model mis-specification; invalid distances/similarities; and incorrect likelihood forms
Independence of measurements/observations (unless explicitly modelled otherwise)	Inflated significance; underestimated uncertainty; and misleading confidence/credible intervals
Appropriate scale and transformation of variables whenever scale-sensitive methods are used	Distorted distances; unreliable predictions; and dominance of high-variance features
Absence of severe outliers unless method is robust or outliers are explicitly modelled.	Skewed parameters; cluster distortion; unstable fits; and misleading trends
Appropriate sample size for the chosen analytical technique (in the absence of a formal power calculation)	Unstable estimates; wide uncertainty; high variance; and unreliable re-sampling or simulation.
Correct temporal ordering for time-dependent (and causal inference) analyses	Spurious autocorrelation; invalid forecasting; and causality misinterpretation
Appropriate randomness assumptions when either re-sampling or simulation is used	Invalid bootstrap intervals; biased Monte Carlo outputs; and incorrect uncertainty estimates

Correct specification of the objective function or loss criterion	Model optimises the wrong behaviour; invalid conclusions; and poor predictive performance
---	---

Table A2.2. A non-exhaustive list of the particular (technique-specific) parametric/non-parametric assumptions that need to apply when using each of the 11 techniques; and the potential analytical and inferential consequences were any of these assumptions not to hold.

Technique	Parametric/non-parametric assumptions	Consequences were the assumptions not to hold
Linear Regression	Linearity of relationships; additive effects; homoscedastic residuals; normally distributed residuals for exact inference; no multicollinearity; and independent errors	Biased coefficients; invalid standard errors; inflated Type I errors; unstable estimates; and misleading inference
Logistic Regression	Correct link function (logit typically being appropriate); linearity of log-odds in predictors; absence of complete separation; independent errors; and correct distributional form (i.e. Bernoulli responses)	Infinite coefficients; non-convergence; mis-calibrated predicted probabilities; and biased odds ratios
K-Nearest Neighbours	Meaningful distance metric; local smoothness (similarity of nearby points); absence of irrelevant or dominating features; and balanced class structure for classification	Distorted neighbour sets; high variance or high bias; poor classification/regression performance; and meaningless similarity structure
K-Means Clustering	Cluster shapes approximately spherical or convex; clusters separable via Euclidean distance; similar cluster variance; and meaningful centroid representation	Mis-clustering; merged or fragmented clusters; unstable solutions; and misleading group interpretations
ANCOVA (Analysis of Covariance)	Homogeneity of regression slopes; linear relationship between covariate and outcome; correct specification of covariates; independent residuals; and homoscedasticity	Biased adjusted means; invalid comparisons across groups; and incorrect significance tests
Trend Plotting	Underlying process is smooth enough for chosen smoother technique; independence of, or correctly modelled, dependence; appropriateness of smoothing window; and absence of structural breaks (unless modelled)	False trend detection; noise mistaken for signal; significant structure masking; and misleading visual inference
ARIMA (Autoregressive Integrated Moving Average)	Stationarity (or stationarity achieved through differencing); invertibility; correct autoregressive and moving-average order; residual independence; homoscedastic residuals; and absence of unmodelled structural breaks	Poor forecasts; spurious autocorrelation; biased parameter estimates; and unstable time-series behaviour
Bayesian Analysis	Correct likelihood form; coherent specification of prior; compatibility of prior and likelihood; sufficient MCMC (Markov Chain Monte Carlo) convergence; and <i>a posteriori</i> integrability	Domination of/by prior; misleading <i>a posteriori</i> findings/inferences; invalid credible intervals; non-converged chains; and incorrect uncertainty estimation
Monte Carlo Simulation	Correct input distributions; valid dependence and correlation structure; sufficient simulation size; and integrity of RNG (random number generator)	Inaccurate risk/uncertainty estimates; biased outputs; unstable results; and simulation artefacts

Cross-Validation	Independence across folds; correct fold structure (e.g., temporal blocking for time series); representativeness of training and testing sets; and a consistent evaluation metric	Inflated or deflated performance estimates; invalid model selection; leakage; and misleading generalisability
Bootstrapping	Sample approximates population distribution; independence or correct block design for time series; sufficient re-sample size; and correct bootstrap type (i.e. parametric, non-parametric or block).	Misleading confidence intervals; incorrect/wrong coverage; underestimated variance; and invalid inference for dependent data

Table A3.1 A non-exhaustive list of the generic (technique-agnostic) sources of non-systematic error (imprecision) and systematic error (bias) that can affect the accuracy, precision and interpretation of findings generated using any of the 11 techniques (and *many/most* other quantitative analytical/statistical techniques); and the analytical and inferential consequences were steps not to have been taken so as to avoid, attenuate, mitigate, or acknowledged and accommodate these sources of imprecision/bias (and their effects).

Sources of imprecision and bias	Consequences were imprecision and bias not addressed
Sampling error (random fluctuations with respect to a finite sample)	Low precision; wide intervals; unstable estimates; and misinterpretation of random variation as if it reflected meaningful structure
Sampling bias (non-representative samples/selection bias)	Systematic mis-estimation; invalid generalisability; and biased predictions and inference
Measurement error (random noise in presentation of variable[s] and/or their measurement/observation)	Attenuated relationships; reduced power; increased variance; and masked effects
Measurement bias (systematic under/over recording or reporting, and misclassification)	Systematic distortion of parameters; biased coefficients; and incorrect inference
Confounding (unmeasured variables causing both exposure/cause and outcome, or focal predictor and target)	Spurious associations; misleading causal interpretation; and biased parameter estimates
Model mis-specification (wrong functional form; omitted interactions; and/or incorrect structural assumptions)	Bias; residual structure; invalid inference; and misleading conclusions
Data preprocessing bias (improper scaling; transformations; and/or filtering)	Distorted distances; incorrect trend patterns; and unreliable model behaviour
Outlier influence (extreme values not accounted for)	Parameter instability; misleading clusters; distorted regression lines; and invalid forecasts
Dependence structure mismanagement (unmodelled autocorrelation, clustering, or grouping)	Underestimated uncertainty; inflated Type I errors; and false significance
Information leakage (test data contaminating the training procedure)	Severely inflated performance estimates; and invalid generalisability
Algorithmic instability (sensitivity to random initialisation, or to small perturbations)	Low reliability; non-reproducible results; and unstable inferences
Hyperparameter/tuning bias (whether tuned on test set or over-optimised)	Overfitting; unrealistic performance; and degraded real-world accuracy
Over-smoothing or under-smoothing (for time series or trend analyses)	Masked structure or exaggerated noise; and incorrect conclusions about trends or cycles

Simulation or re-sampling randomness (Monte Carlo or bootstrap variability)	Unstable uncertainty estimates; and misleading intervals (if insufficient samples drawn)
Interpretation bias (misreading/misinterpretation of statistical or probabilistic outputs)	Incorrect substantive conclusions; and miscommunication of risk, effect sizes, or uncertainty

Table A3.2 A non-exhaustive list of the particular (technique-specific) sources of non-systematic error (imprecision) and systematic error (bias) that can affect the accuracy, precision and interpretation of findings generated using each of the 11 techniques; and the analytical and inferential consequences were steps not to have been taken so as to avoid, attenuate, mitigate, or acknowledged and accommodate these sources of imprecision/bias (and their effects).

Technique	Sources of imprecision and bias	Consequences were imprecision and bias not addressed
Linear Regression	Heteroscedastic residuals; multicollinearity; influential outliers; and omitted nonlinear terms	Biased coefficients; incorrect standard errors; unstable estimates; and misleading inference
Logistic Regression	Complete/quasi separation; rare events bias; mis-calibrated class imbalance; and mis-specified link	Infinite or unstable coefficients; biased odds ratios; and poor probability calibration
K-Nearest Neighbours	'Curse of dimensionality'; unscaled variables; irrelevant features; and class imbalance	Distorted neighbour sets; poor classification; and either high variance or high bias
K-Means Clustering	Poor centroid initialisation; non-spherical clusters; sensitivity to scaling; and empty clusters	Incorrect cluster assignment; unstable clustering; and misleading segmentation
ANCOVA (Analysis of Covariance)	Violation of slope homogeneity; mis-specified covariates; and imbalance across groups	Biased adjusted means; invalid comparisons; and inflated Type I errors.
Trend Plotting	Incorrect smoothing span; unremoved seasonality; and failure to handle structural breaks	False trends; masked patterns; and misinterpretation of time-dependent behaviour
ARIMA (Autoregressive Integrated Moving Average)	Incorrect differencing order; unmodelled seasonality; residual autocorrelation; and parameter non-invertibility	Spurious autocorrelation; biased forecasts; instability; and misleading time-series structure
Bayesian Analysis	Poorly chosen <i>priors</i> ; conflict between prior and likelihood; non-converged MCMC (Markov Chain Monte Carlo); and autocorrelated chains	Prior domination; incorrect <i>a posteriori</i> ; invalid credible intervals; and unreliable inference
Monte Carlo Simulation	Incorrect input distributions; mis-specified correlations; insufficient iterations; random number generator deficiencies	Invalid uncertainty quantification; biased risk estimates; and simulation artefacts
Cross-Validation	Improper fold assignment; temporal leakage; stratification failure; and inconsistent evaluation metric	Inflated/deflated performance; wrong model selection; and invalid generalisability
Bootstrapping	Incorrect variance estimates; wrong interval coverage; and misleading uncertainty assessments	Incorrect variance estimates; wrong interval coverage; and misleading uncertainty assessments

Table A4. 1 A non-exhaustive list of the generic (technique-agnostic) diagnostic assurance measures that can be used for all 11 of the techniques (and *many/most* other quantitative analytical/statistical techniques) to assess whether: the most appropriate analytical choices/decisions have been made (see Table A1.1-A1.3, above); all necessary assumptions apply (see Table A2.1-A2.2, above); and all potential sources of imprecision and bias have been addressed (see Table A3.1-A3.2, above).

Diagnostic assurance measures available	Consequences should diagnostic assurance measures not be applied (and reported)
Explicitly defining the target population and the estimand (the estimate sought/desired) ^{1,2}	Misaligned analyses; results that answer the 'wrong' question; and non-transportable conclusions
Pre-registering the analysis plan or 'locking' a protocol ^{1,3}	Researcher-adjusted degrees of freedom; p-hacking; inflated Type I error rate; and overfitted narratives
Auditing: data provenance, missingness, and measurement error ^{4,5}	Biased estimates; spurious associations; unknown and unacknowledged uncertainty; and selective omission
Conducting train/validation/test separation and leakage checks ⁶	Overstated performance; failures of generalization; and misleading model comparisons
Conducting exploratory data analysis (to assess distributions, outliers, and scale) ^{4,7}	Model misfit with leverage points dominating; and/or invalid standard errors
Undertaking technique-relevant assumption mapping (e.g. regarding: linearity; independence; and stationarity) ^{7,8}	Hidden violations that degrade validity; and biased/inefficient or inconsistent estimates
Ensuring analytical features are engineered with appropriate scaling or encoding ⁷	Distance-based/generalised linear model fits that are distorted; non-convergence; and unstable coefficients
Developing and applying a confounding control strategy (through design or using a DAG-based adjustment set) ^{4,9}	Invalid causal claims; 'omitted-variable bias'; and collider bias (if these are incorrectly conditioned on)
Adopting a single, pre-specified model cross-validation or selection/evaluation rule to avoid <i>post hoc</i> cherry-picking ⁶	Overfitting/underfitting; cherry-picked metrics; and irreproducible choices
Conducting diagnostic residuals checks and influence analysis ^{4,7}	Undetected heteroscedasticity, autocorrelation, and/or nonlinearity; with results driven by a few data points
Ensuring standard errors are robust; or conducting variance modelling when needed ⁷	Invalid confidence intervals and p-values; leading to inflated false positives and/or false negatives
Controlling for multiple testing/error-rate (e.g. Family-Wise Error Rate or False Discovery Rate) ⁷	Excessive numbers of false discoveries; and an unreliable portfolio of 'insights'
Conducting sensitivity analyses (to evaluate specifications, <i>priors</i> , and bandwidths) ^{4,9,10}	Fragile conclusions; and undisclosed model dependencies
Undertaking uncertainty quantification (e.g. by generating confidence, credible and predictive intervals) ^{1,8}	Overconfident claims; and unassessed risks
Conducting external validation or out-of-time validation ⁶	Non-transportable models; and enhanced risk of 'surprise failure' following model deployment
Evaluating reproducibility (of code, seeds, and versions) ³	Irreplicable findings; and undiagnosable discrepancies

Applying transparent reporting (including of: analysis limitations, assumptions and data bounds) ²	Misinterpretation by customers/consumers; and potential misuse/misapplication in decision-making
---	--

¹Nance et al. (2024); ²Sterne et al. (2016); ³Nosek et al. (2018); ⁴Yu et al. (2024); ⁵Little and Rubin (2019); ⁶Varma and Simon (2006); ⁷Fox (2015); ⁸Gelman et al. (2020); ⁹Textor et al. (2016); ¹⁰VanderWeele and Ding (2017).

Table A4.2 (part 1 of 2) A non-exhaustive list of the particular (technique-specific) diagnostic assurance measures that can be used for each of the 11 of the techniques to assess whether: the most appropriate analytical choices/decisions have been made (Table A1.1-A1.3, above); all necessary assumptions apply (Table A2.1-A2.2, above); and all potential sources of imprecision and bias have been addressed (Table A3.1-A3.2, above).

Technique	Diagnostic assurance measures available	Consequences should diagnostic assurance measures not be applied (or reported)
Linear Regression	Residual plots for linearity/normality; heteroscedasticity tests (Breusch–Pagan); robust/heteroscedasticity-consistent standard errors; multicollinearity (variance inflation factor [VIF]); ‘influence’ (Cook’s D); specification tests; and interaction checks ^{1,2}	Biased/inefficient estimates; invalid standard errors/confidence intervals; spurious significance from nonlinearity; and results driven by just a few points
Logistic Regression	Linearity in the logit (Box–Tidwell test); separation checks and penalization (Firth bias-reduced logistic regression or Ridge/L2-penalised logistic regression); calibration (reliability plots, Brier scores, and/or the Hosmer–Lemeshow goodness-of-fit test); measures of class imbalance handling, multicollinearity and threshold-independent metrics (area under the curve [AUC] ROC, and/or precision-recall curve [PRC]) ^{3,4,5}	Poor calibration; unstable/biased coefficients; misleading accuracy as a result of imbalance; and odds ratios misinterpreted as risk ratios
K-Nearest Neighbours	Feature scaling; ‘k’ chosen via cross-validation; selection of appropriate distance metric; class imbalance handling; leakage prevention (temporal splits if needed); and reduction of dimensionality if high-dimensional data (high-D) ^{6,7}	Noisy, unstable predictions; optimism from leakage; and distance that is meaningless in high dimensions
K-Means Clustering	Scale features; multiple random initializations; choice of ‘k’ (silhouette, gap statistic); assess cluster stability; handle outliers; and interpret centroids ^{8,9}	Arbitrary or non-replicable clusters; misleading “natural groups”; and spurious downstream profiles
ANCOVA (Analysis of Covariance)	Verify homogeneity of regression slopes (parallel slopes); residual diagnostics; covariate reliability; multiple-comparison adjustments; and balance/randomization checks ^{2,10}	Invalid adjusted means; biased group effects; inflated Type I error rate; and misattribution to baseline

Trend Plotting	Choose smoothing (locally estimated scatterplot smoothing [LOESS]; or moving average [MA]) with rationale; preserve time order; show uncertainty bands; adjust or model temporal periodicity (if relevant); and test for change points ^{11,12}	Over-reading noise as if it were a trend; seasonality overlooked/omitted; and false claims of structural change
-----------------------	---	---

Derived from the following sources: ¹Breusch and Pagan (1979); ²Fox (2015); ³Hosmer et al. (2013); ⁴Firth (1993); ⁵Steyerberg et al. (2010); ⁶Cover and Hart (1967); ⁷Aggarwal et al. (2001); ⁸Kaufman and Rousseeuw (2005); ⁹Tibshirani et al. (2001); ¹⁰Maxwell and Delaney (2004); ¹¹Cleveland (1979); and ¹²Killick et al. (2012).

Table A4.2 (part 2 of 2) A non-exhaustive list of the particular (technique-specific) diagnostic assurance measures that can be used for each of the 11 of the techniques to assess whether: the most appropriate analytical choices/decisions have been made (Table A1.1-A1.3, above); all necessary assumptions apply (Table A2.1-A2.2, above); and all potential sources of imprecision and bias have been addressed (Table A3.1-A3.2, above).

Technique	Diagnostic assurance measures available	Consequences should diagnostic assurance measures not be applied (or reported)
ARIMA (Autoregressive Integrated Moving Average)	Check stationarity (augmented Dickey Fuller test/Kwiatkowski–Phillips–Schmidt–Shin test); identify ARIMA orders: (p) autoregressive, (d) differencing, and (q) moving-average) via autocorrelation function (AF), partial autocorrelation function (PACF) and information criteria (IC); ensure residual whiteness/white noise; include seasonal ARIMA (SARIMA) when appropriate; run stability checks; and evaluate out-of-sample forecasts/intervals ^{13,14}	Mis-specified dynamics; autocorrelated residuals; biased forecasts; and underestimated intervals
Bayesian Analysis	Specify/justify priors; diagnose Markov chain Monte Carlo (MCMC) convergence (R-hat, effective sample size (ESS), and trace plots); posterior predictive checks; prior/posterior sensitivity; and identifiability review ^{15,16}	Non-converged or prior-dominated <i>a posteriori</i> ; misleading credible intervals; and undetected misfit
Monte Carlo Simulation	Validate input distributions and dependencies; justify scenario design; sufficient runs to stabilize estimates; consider variance reduction; and sanity-check vs. historical data ^{17,18}	Unreliable scenario probabilities; spurious tail risks (or missed extremes); and unstable decision metrics
Cross-Validation	Proper fold construction (stratified, group, or time-series cross-validation [CV]); prevent leakage; fix evaluation metric; report mean/variance across folds; and repeated CV for stability ^{19,20}	Optimistic performance; non-generalizing/generalisable models; and unstable model selection

Bootstrapping	Resample with replacement; choose appropriate bootstrap for dependence (block/cluster); use sufficient replicates; use percentile/bias-corrected and accelerated (BCa) intervals; and check interval stability ^{21,22}	Under/overestimated uncertainty; invalid confidence intervals with dependencies; and overconfident claims of robustness
---------------	---	---

Derived from the following sources: ¹³Box et al. (2015); ¹⁴Hyndman and Athanasopoulos (2021);); ¹⁵Gelman et al. (2020); ¹⁶Vehtari et al. (2021); ¹⁷Glasserman (2004); ¹⁸Helton and Davis (2003); ¹⁹Varma and Simon (2006); ²⁰Bergmeir et al. (2018); ²¹Efron and Tibshirani (1993); and ²²Davison and Hinkley (1997)

References

- Aggarwal CC, Hinneburg A, Keim DA. On the surprising behavior of distance metrics in high dimensional space. In: Van den Bussche J and Vianu V (eds.) *Lecture Notes in Computer Science* 2001; Berlin, Heiledberg (D): Springer. pp. 420-34. DOI: 10.1007/3-540-44503-X_27
- Agresti A. *Foundations of Linear and Generalized Linear Models*, 1st Edition. Hoboken (NJ, USA): Wiley; 2015. 472pp. ISBN: 9781118730034
- Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis* 2018; 120: 70-83. DOI: 10.1016/j.csda.2017.11.003
- Bishop CM. *Pattern Recognition and Machine Learning*, 1st Edition. New York (NY, USA): Springer; 2006. 738pp. ISBN: 9780387310732
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time Series Analysis: Forecasting and Control*, 5th Edition. Hoboken (NJ, USA): Wiley; 2015. 712pp. ISBN: 9781118675021
- Box GEP. Science and statistics. *Journal of the American Statistical Association* 1976; 71: 791-799. DOI: 10.1080/01621459.1976.10480949
- Box GEP. Robustness in the strategy of scientific model building, In: Launer RL, Wilkinson GN (eds.) *Robustness in Statistics*. New York (NY, USA): Academic Press; 1979. pp. 201–236. DOI: 10.1016/B978-0-12-438150-6.50018-2
- Breusch TS, Pagan AR. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* 1979; 47:1287–94. DOI: 10.2307/1911963
- British Army. *Operator Military Intelligence*. British Army Find a Role Website. Archived [here](#) on 30MAY26.
- Casella G, Berger RL. *Statistical Inference*, 2nd Edition. Pacific Grove (CA, USA): Duxbury Press; 2002. 660pp. ISBN: 9780534243128

Christensen R, Ranstam J, Overgaard S, Wagner P. Guidelines for a structured manuscript: Statistical methods and reporting in biomedical research journals. *Acta Orthopaedica* 2023; 94: 243-9. DOI: 10.2340/17453674.2023.11656

Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; 74: 829-36. DOI: 10.1080/01621459.1979.10481038

Corera G. How Israel builds its hi-tech start-ups. Technology, BBC News website 2016; 14 Oct. Archived [here](#) on 14OCT16.

Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory* 1967; 13: 21-7. DOI: 10.1109/TIT.1967.1053964

Davison AC, Hinkley DV. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press; 1997. 582pp. ISBN: 9780511802843

Duffield J. Statistical analytical techniques for intelligence analysis. *RUSI Journal* 2026; 171: 1-12. DOI: 10.1080/03071847.2026.2646065

Efron B, Hastie T. *Computer Age Statistical Inference*, 1st Edition. Cambridge (UK): Cambridge University Press; 2016. 495pp. ISBN: 9781107149899

Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*, 1st Edition. New York (NY): Chapman and Hall/CRC; 1993. 456pp. ISBN: 978-0412042317

Ellison GTH. Using directed acyclic graphs (DAGs) to represent the data generating mechanisms of disease and healthcare pathways: A guide for educators, students, practitioners and researchers. In: Farnell DJJ, Medeiros Mirra R (eds). *Teaching Biostatistics in Medicine and Allied Health Sciences*. Cham (Switzerland): Springer; 2023. pp. 61-101. DOI: 10.1007/978-3-031-26010-0_6

Ellison GTH. Statistics in Intelligence Analysis: "a little learning is a dang'rous thing". *RUSI Journal* 2026; DOI: 0.1080/03071847.2026.2683229

Elwert F, Winship C. Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology* 2014; 40: 31-53. DOI: 10.1146/annurev-soc-071913-043455

Firth D. Bias reduction of maximum likelihood estimates. *Biometrika* 1993; 80: 27-38. DOI: 10.1093/biomet/80.1.27

Fox J. *Applied Regression Analysis and Generalized Linear Models*, 4th Edition. Los Angeles (CA): SAGE Publications; 2015. 824pp. ISBN: 978-1452205663

Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P, Modrák M. Bayesian workflow. *arXiv* 2020; 03 Nov: 1-77. DOI: 10.48550/arXiv.2011.01808

Glasserman P. *Monte Carlo Methods in Financial Engineering*. New York (NY): Springer; 2003. 596pp. ISBN: 978-0387004518

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*, 3rd Edition. Boca Raton (FL, USA): CRC Press; 2014. 662pp. ISBN: 9781439840955

Goodfellow I, Bengio Y, Courville A. *Deep Learning*, 1st Edition. Cambridge (MA, USA): MIT Press; 2016. 800pp. ISBN: 9780262035613

Hardwicke TE, Salholz-Hillel M, Malički M, Szűcs D, Bendixen T, Ioannidis JPA. (2023) Statistical guidance to authors at top ranked journals across scientific disciplines. *American Statistician* 2023; 77: 239-47 DOI: 10.1080/00031305.2022.2143897

Harrell FE. *Regression Modeling Strategies*, 2nd Edition. Cham (Switzerland): Springer; 2015. 582pp. ISBN: 9783319194240

Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*, 2nd Edition. New York (NY, USA): Springer; 2009. 745pp. ISBN: 9780387848570

Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering and System Safety* 2003; 81: 23-69. DOI: 10.1016/S0951-8320(03)00058-9

Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*, 3rd Edition. Hoboken (NJ): Wiley; 2013. 510pp. ISBN: 9780470582473

Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and Practice*, 3rd Edition. Melbourne, Australia: OTexts; 2021. 442pp. ISBN: 9780987507136

Ioannidis, JPA. Statistical biases in science communication: What we know about them and how they can be addressed. In: Jamieson KH, Kahan DM, Scheufele DA (eds). *The Oxford Handbook of the Science of Science Communication* (Online Edition). Oxford Academic, Oxford Library of Psychology: Oxford (UK) 2017; 2017; 06 Jun. DOI: 10.1093/oxfordhb/9780190497620.013.11

Ioannidis, JPA. What have we (not) learnt from millions of scientific papers with p values? *American Statistician* 2019; 73 (Suppl 1): 20-5. DOI: 10.1080/00031305.2018.1447512

James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*, 2nd Edition. New York (NY, USA): Springer; 2021. 612pp. ISBN: 9781071614174

Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken (NJ): Wiley; 1990. 342pp. ISBN: 9780471878766

Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* 2012; 107: 1590-8. DOI: 10.1080/01621459.2012.737745

Kim J, Kim DH, Kwak SG. Comprehensive guidelines for appropriate statistical analysis methods in research. *Korean Journal of Anesthesiology* 2024; 77: 503-17. DOI: 10.4097/kja.24016

Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*, 5th Edition. Boston (MA, USA): McGraw-Hill; 2005. 1396pp. ISBN: 9780073108742.

Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. *International Journal of Nursing Studies* 2015; 52: 5-9. DOI: 10.1016/j.ijnurstu.2014.09.006

Le Carre J. *Tinker, Tailor, Sailor, Spy*. London (UK): Hodder and Stoughton; 1974. 355pp. ISBN: 0-340-18879-0

Little RJA, Rubin DB. *Statistical Analysis with Missing Data*, 3rd Edition. Hoboken (NJ): Wiley; 2019. 462pp. ISBN: 978-0-470-52679-8

Maxwell SE, Delaney HD. *Designing Experiments and Analyzing Data: a Model Comparison Perspective*, 2nd Edition. New York (NY): Routledge; 2003. 920pp. ISBN: 9781410609243

Montgomery C, Engelmann L. Epidemiological publics? On the domestication of modelling in the era of COVID-19. *Somatosphere* 2020; 10 Apr. Archived [here](#) on 02MAR24.

McElreath R. *Statistical Rethinking*, 2nd Edition. Boca Raton (FL, USA): CRC Press; 2020. 562pp. ISBN: 9780367139919.

MOD (Ministry of Defence). *Intelligence, Counter-intelligence and Security Support to Joint Operations*. Joint Doctrine Publication JDP 2-00, 4th Edition – Development, Concepts and Doctrine Centre, Ministry of Defence, UK; 2023: 181pp. Archived [here](#) on 20NOV23.

MOD (Ministry of Defence). *Cyber and Electromagnetic Activities*. Joint Doctrine Note JDN 1/18, 1st Edition – Development, Concepts and Doctrine Centre, Ministry of Defence, UK; 2018: 54pp. Archived [here](#) on 27JUN24.

Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*, 6th Edition. Hoboken (NJ, USA): Wiley; 2021. 656pp. ISBN: 9781119722106

Murphy KP. *Machine Learning: A Probabilistic Perspective*, 1st Edition. Cambridge (MA, USA): MIT Press; 2012. 1067pp. ISBN: 9780262018029

Nance N, Petersen ML, van der Laan M, Balzer LB. The causal roadmap and simulations to improve the rigor and reproducibility of real-data applications. *Epidemiology* 2024; 35: 791-800. DOI: 10.1097/EDE.0000000000001773

Nosek BA, Ebersole CR, DeHaven AC, Mellor DT. The preregistration revolution. *Proceedings of the National Academy of Sciences of the USA*. 2018; 115: 2600–6. DOI: 10.1073/pnas.1708274114

OxfordAQA (Oxford Assessments and Qualifications Alliance). *GCSE Mathematics 8300 – Specification*. OxfordAQA Website; 2026. Archived [here](#) on 07OCT24.

Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, 1st Edition. New York (NY, USA): Wiley; 1987. 258pp. ISBN: 9780471081934

RAF (Royal Air Force). *Intelligence*. Royal Airforce Recruitment Website. Archived [here](#) on 01DEC25.

RN (Royal Navy). 2026a. *Warfare Rating*. Royal Navy Careers Website. Archived [here](#) on 25FEB26.

RN (Royal Navy). 2026b. *Defence Aptitude Assessment*. Royal Navy Careers Website 2026; Archived [here](#) on 28MAY26.

Shalizi CR. *Advanced Data Analysis from an Elementary Point of View*, 1st Edition. Cambridge (UK): Cambridge University Press; 2021. 736pp. ISBN: 9781107190204

Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *British Medical Journal* 2016; 355: i4919. DOI: 10.1136/bmj.i4919

Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattan MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; 21: 128-38. DOI: 10.1097/EDE.0b013e3181c30fb2

Thomas Z, Kruppa M. This Israeli Army Unit Has Become an Incubator for Tech Startups. Tech News Briefing, Wall Street Journal: New York (NY) 2024; 02 Sep. Archived [here](#) on 03SEP24.

Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GTH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* 2016; 45: 1887-1894. DOI: 10.1093/ije/dyw341

Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society B* 2001; 63: 411-23. DOI: 10.1111/1467-9868.00293

VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine* 2017; 167: 268-74. DOI: 10.7326/M16-2607

Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 2006; 7: 91. DOI: 10.1186/1471-2105-7-91

Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-normalization, folding, and localization: an improved R-hat for assessing convergence of MCMC. *Bayesian Analysis* 2021;16: 667-718. DOI: 10.1214/20-BA1221

Yu X, Zoh RS, Fluharty DA, Mestre LM, Valdez D, Tekwe CD, Vorland CJ, Jamshidi-Naeini Y, Chiou SH, Lartey ST, Allison DB. Misstatements, misperceptions, and mistakes in

controlling for covariates in observational research. *eLife* 2024; 13: e82268. DOI: 10.7554/eLife.82268

Yuval, A. 'Lavender': The AI machine directing Israel's bombing spree in Gaza. +972 Magazine 2024; 03 Apr. Archived [here](#) on 03APR24.

Wasserman L. *All of Statistics*, 1st Edition. New York (NY, USA): Springer; 2004. 440pp. ISBN: 9780387402727

Wasserman L. *All of Nonparametric Statistics*, 1st Edition. New York (NY, USA): Springer; 2006. 268pp. ISBN: 9780387251455

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.