

# Gold Standards in Pharmacovigilance

## The Use of Definitive Anecdotal Reports of Adverse Drug Reactions as Pure Gold and High-Grade Ore

Manfred Hauben<sup>1,2,3,4</sup> and Jeffrey K. Aronson<sup>5</sup>

- 1 Risk Management Strategy, Pfizer Inc., New York, New York, USA
- 2 Department of Medicine, New York University School of Medicine, New York, New York, USA
- 3 Department of Community and Preventive Medicine, New York Medical College, Valhalla, New York, USA
- 4 Department of Pharmacology, New York Medical College, Valhalla, New York, USA
- 5 Department of Clinical Pharmacology, Radcliffe Infirmary, Oxford, England

### Abstract

Anecdotal reports of adverse drug reactions are generally regarded as being of poor evidential quality. This is especially relevant for postmarketing drug safety surveillance, which relies heavily on spontaneous anecdotal reports. The numerous limitations of spontaneous reports cannot be overemphasised, but there is another side to the story: these datasets also contain anecdotal reports that can be considered to describe definitive adverse reactions, without the need for further formal verification. We have previously defined four categories of such adverse reactions: (i) extracellular or intracellular tissue deposition of the drug or a metabolite; (ii) a specific anatomical location or pattern of injury; (iii) physiological dysfunction or direct tissue damage demonstrable by physicochemical testing; and (iv) infection, as a result of the administration of an infective agent as the therapeutic substance or because of demonstrable contamination. In this article, we discuss the implications of these definitive ('between-the-eyes') adverse effects for pharmacovigilance.

Anecdotal reports are generally regarded as being of poor evidential quality. However, we have previously proposed that some anecdotal reports of adverse drug reactions are underappreciated sources of definitive associations.<sup>[1]</sup> In this article, we extend our discussion to a detailed consideration of the relevance to pharmacovigilance of such reactions, which we propose can serve as perfect gold standards, or at least very high-grade ore.

We first reiterate briefly our previous observation, namely that there are at least four categories of adverse events that can provide definitive evidence

of true drug-event associations, without the need for formal validation. The four groups, examples of which are listed in table I, are as follows:

1. Extracellular or intracellular tissue deposition of the drug or a metabolite.
2. A specific anatomical location or pattern of injury.
3. Physiological dysfunction or direct tissue damage demonstrable by physicochemical testing.
4. Infection, as a result of the administration of an infective agent as the therapeutic substance or because of demonstrable contamination.

**Table I.** Definitive anecdotal adverse drug reactions

Event	Drug examples	Confirmatory tests/characteristics
<b>1a. Extracellular deposition of drug or metabolite</b>		
Biliary lithiasis or pseudolithiasis	Ceftriaxone, <sup>[2]</sup> sulindac <sup>[3,4]</sup>	Infrared spectroscopy
Nephrolithiasis	Aciclovir, amoxicillin, ciprofloxacin, ephedrine/guaifenesin, felbamate, indinavir, magnesium trisilicate, methotrexate, primidone, sulfasalazine, sulfonamides, triamterene <sup>[5,6]</sup>	Infrared spectroscopy, x-ray diffraction, mass spectroscopy
Conjunctival cysts	Tetracycline <sup>[7,8]</sup>	Wood's lamp
Exogenous lipoid pneumonia	Mineral oil <sup>[9]</sup>	Gas chromatography/mass spectrometry
Pharmacobezoars	Colestyramine, sucralfate, modified-release formulations, guar gum, ion exchange resins <sup>[10]</sup>	Visual inspection
Baroliths	Barium <sup>[11]</sup>	X-ray, visual inspection
<b>1b. Intracellular deposition of drug or metabolite</b>		
Corneal microprecipitates	Ciprofloxacin <sup>[12]</sup> Gold <sup>[13]</sup>	Scanning EM, HPLC Confocal microscopy
Crystalline retinopathy	Methoxyflurane, <sup>[14]</sup> canthaxanthin <sup>[15]</sup>	HPLC
Crystal-storing histiocytosis	Aluminium-containing vaccines <sup>[16]</sup> Clofazimine <sup>[17]</sup>	Electron microprobe analysis Visual inspection, polarising microscopy
Intraglomerular crystal deposition	Foscarnet <sup>[18]</sup>	Fourier transform infrared spectroscopy
Lymphadenopathy	Gold <sup>[19]</sup>	Light microscopy, scanning EM
Deposition in nails or lunulae	Tetracycline <sup>[20]</sup> Clofazimine <sup>[21]</sup>	Wood's lamp Light microscopy
Skin pigmentation	Amiodarone <sup>[22]</sup>	HPLC, EM, energy dispersive x-ray microanalysis
<b>2. Specific anatomical location or pattern of injury</b>		
Esophageal ulcers	Bisphosphonates, potassium chloride, quinidine, tetracyclines <sup>[23]</sup>	Localisation to areas of oesophageal lesions
Extravasation reactions	Cancer chemotherapeutic agents <sup>[24]</sup>	Anatomical contiguity to drug administration
Fulminant encephalomyelitis	Inadvertent intrathecal ionic contrast medium, <sup>[25]</sup> inadvertent intrathecal vincristine <sup>[26]</sup>	Anatomical pattern of injury
Nodulosis	Apomorphine <sup>[27]</sup>	Anatomical contiguity to drug administration
Oral damage	Topical oral salicylates, <sup>[28]</sup> topical ecstasy <sup>[29]</sup>	Application site localisation
Nasal damage	Topical cocaine <sup>[30]</sup>	Application site localisation
Haemangiosarcoma	Thorotrast <sup>[31]</sup>	Anatomical localisation in sites of drug accumulation or persistence
<b>3. Physicochemical dysfunction or tissue damage</b>		
Oligohidrosis	Topiramate <sup>[32]</sup> Zonisamide <sup>[33,34]</sup>	Iontophoresis Acetylcholine loading test, heat-loading test
Photosensitivity	Carbamazepine, dapson, fenofibrate, flutamide, certain NSAIDs, triflusa <sup>[35-37]</sup>	Phototesting, photopatch testing
Taste disturbance	Certain NSAIDs <sup>[38]</sup>	Gustatometry, electrogustatometry
Dry mouth	Omeprazole <sup>[39]</sup>	Measurement of salivary flow

*Continued next page*

Table I. Contd

Event	Drug examples	Confirmatory tests/characteristics
<b>4. Infection-related</b>		
Sepsis unrelated to product contamination	Bacille Calmette-Guérin, <sup>[40]</sup> lactobacillus, <sup>[41]</sup> mumps vaccine <sup>[42]</sup>	Polymerase chain reaction, DNA enzyme immunoassay electrophoresis, bacterial culture, strain typing, DNA fingerprinting
Sepsis due to product contamination	Intravenous gentamicin, <sup>[43]</sup> propofol <sup>[44]</sup>	Endotoxin assay, plasmid and restriction- endonuclease analysis

EM = electron microscopy; HPLC = high-performance liquid chromatography.

We have colloquially called these reactions ‘between-the-eyes’ reactions, because the diagnosis is obvious, or almost so, in the individual patient. In some cases, the diagnostic value of the event can be enhanced by further investigation, but the conclusion will always be related to the individual affected.

This gives a new perspective on anecdotal reports of adverse drug reactions, predicated on concepts that are not included in traditional causality assessment procedures, which tend to emphasise chronological characteristics of reported associations (i.e. the time courses of challenge, dechallenge and rechallenge).<sup>[45,46]</sup> We intend no criticism of these traditional criteria, which can provide good evidence for causation, for example in well documented cases of immediate (e.g. ‘end-of-the-needle’) events, in which the temporal relation is so striking in otherwise stable circumstances that alternative causes can be excluded.

We begin with a general discussion of gold standards in pharmacovigilance and other biomedical settings. We then discuss the need for gold standards in pharmacovigilance and finally the availability of such gold standards.

### 1. Thesis: the Demand for Gold Standards in Biomedical Science

There are many publications on gold standards in the context of medical screening, diagnostic tests and epidemiological studies. The use of an imperfect gold standard can introduce bias into the calculation of standard performance parameters (for example, sensitivity, specificity, predictive value and receiver operating characteristic curves).<sup>[47-49]</sup> The significance and direction of such bias depends on multiple factors, including whether the new test and

the reference standard are ‘conditionally independent’, the quality of the imperfect gold standard (that is, how well the imperfect gold standard correlates with the outcome of interest) and the specific metric being considered.<sup>[47]</sup>

Gold standards can only provide useful performance estimates if they are highly correlated with an outcome of interest and if the research question is framed in appropriate common sense terms.<sup>[48-50]</sup> Clinical scenarios often lack ‘perfect’ or ‘pure’ gold standards, because the target environment or outcome of interest is imprecise.<sup>[48,50,51]</sup> Even assessment of tissue pathology is associated with inter-observer variability.<sup>[52]</sup>

Other approaches can be used to improve estimates derived from proxy gold standards in screening, diagnosis and epidemiological studies, including discrepant analysis, latent class models and Bayesian methods.<sup>[53]</sup> There is a growing body of epidemiological research on making causal inferences from observational data when there are multiple sources of bias and error. These have yet to be specifically applied to pharmacovigilance, but they are worthy of exploration.<sup>[54,55]</sup>

### 2. Antithesis: the Need for Gold Standards in Pharmacovigilance

Pharmacovigilance has a dual nature: the search for ‘signals’ of adverse events that are novel in terms of their nature, severity and/or frequency is accomplished by evaluation of single reports and case series,<sup>[56,57]</sup> as well as numerical approaches akin to ‘anomaly detection’ in other forms of public health surveillance.<sup>[58]</sup> Thus, signal detection and evaluation is a multi-step process that often uses multiple tests and data streams.

Spontaneous reporting systems, composed largely of anecdotal case reports that are not peer reviewed, are currently the cornerstone of postmarketing signal detection. Such databases may be very large, sparse and plagued by data distortion and corruption at the level of the individual report. Furthermore, the overall sampling mechanism reflects a convenience sample, based on differential reporting across drugs and events. In addition, the data lack information on the numbers of patients exposed or at risk.

Causal assessments in pharmacovigilance can occur at the level of both the individual case report and the overall association (i.e. the case series). These are not independent processes, as the amount of information contained in individual cases determines the number of cases required to achieve a critical evidentiary mass for the purposes of decision making; on average, the more definitive the evidence per case (either for or against causality), the lower is the critical evidentiary mass of cases. It may be possible to pin down causally a drug-event association with sufficient certainty with only a few cases if certain causality features are well documented and to implicate the drug with high probability in each case. If the report contains less information and the probability of causation is lower per report, coincidental associations are more likely and more cases will be required to achieve the critical evidentiary mass. This underlines the need for comprehensive reporting of individual cases.<sup>[59]</sup>

With increasingly large databases and massive influxes of reports of adverse events, statistical data-mining algorithms (DMAs) are being more often promoted to assist the process of signal detection in pharmacovigilance. Most contemporary DMAs are variations of 'disproportionality analysis' and can be broadly classified into frequentist types (proportional reporting ratios and reporting odds ratios) and Bayesian types (Bayesian Confidence Propagation Neural Network, and the multi-item gamma Poisson shrinker).<sup>[60]</sup> There is considerable controversy about the value and optimal use of these methods.<sup>[61]</sup> Each has case-dependent strengths and weaknesses,<sup>[62]</sup> and all have significant limitations.<sup>[60]</sup>

Pharmacovigilance DMAs are typically studied by retrospective application to authentic spontaneous reporting systems (although some advocate database simulations as the definitive approach to validation).<sup>[60,63]</sup> Attempts to test, validate and interpret results from these tools are complicated by various factors, both non-methodological, such as potential commercial and intellectual conflicts of interest, and methodological.<sup>[64]</sup> The latter include confirmation bias, the abundance (or perhaps overabundance) of available data-mining options and configurations, the *ad hoc* and subjective nature of threshold selection, the need to assess both 'dynamic' (i.e. 'time-to-signal') and traditional 'static' performance metrics, such as sensitivity, specificity and predictive value, and a claimed lack of gold standards for adjudicating causality in spontaneous reports.<sup>[60,61]</sup> The last problem pertains to decisions about which spontaneously reported adverse events should be included in reference sets for testing DMAs. In real time, the detection of an event precedes its verification. In the current context, event verification is a prerequisite for assessing the procedures and tools that are used in event detection. Reference sets must be constructed from spontaneously reported associations that were unknown at the time of marketing, which can be classified as causal or non-causal with a reasonable degree of confidence (true positive and true negative, respectively), and for which there is a reasonably specific terminology in the adverse effects dictionary used to encode the data.

Causal classification requires a reasonably reliable gold standard for designating a drug-event association as a case or non-case. Here we refer to 'case' versus 'non-case' at the level of the reported drug-event association, rather than the level of individual reports. A purported association may be an ensemble consisting of definitive cases, probable or possible cases, and non-cases. That is to say, the total corpus of reports of a particular drug-event combination may include distinct subpopulations of cases, some demonstrably causal, some causal but not conclusively demonstrated as such, and some not causal at all.

### 3. Synthesis: the Availability of Gold Standards in Pharmacovigilance

Given the data limitations delineated in the previous section, it is not surprising that the claimed lack of gold standards for assessing causality is among the most frequently cited of all obstacles in testing DMAs. Its most extreme expression is the objection to the use of associations for which there is 'no guarantee' of causality. This viewpoint may be of philosophical interest, but it should not be selectively applied to shield emerging methods from scientific scrutiny or to object to findings because they are disagreeable. There are alternative ways to think about case definitions in terms of outcomes of interest that are logically sound and may have greater practical utility in real-world pharmacovigilance scenarios.<sup>[65]</sup> This includes the value of effective surveillance systems in promoting a greater awareness and understanding of complex, dynamic and uncertain environments,<sup>[66]</sup> which in the case of pharmacovigilance includes warning signs of probably causal but unproven associations with potential public health implications.

What constitutes a gold standard for designating cases and non-cases varies according to the circumstances in which it is applied and the objectives of the surveillance system. The range of observed event frequencies in patients taking a drug relative to the corresponding background incidence<sup>[56]</sup> indicates that establishing definite or probable associations (i.e. suitable reference events) across the spectrum of pharmacovigilance scenarios requires analysis of the full range of datasets from case reports, observational studies and randomised clinical trials<sup>[67]</sup> and must exploit both clinical observations and numerical data; this has been termed 'teleoanalysis'.<sup>[59,68]</sup> Nothing is ever irrefutably proven, and even the most pure gold standard is established with a degree of probability, albeit a high one. Therefore, we believe that much of the discussion about absolute or perfect gold standards is too restrictive and demonstrates an under-appreciation of both the uncertainty of the target environment and the range of events of legitimate interest to pharmacovigilance professionals.<sup>[65]</sup>

Thus, immediate (e.g. 'end-of-the-needle') events, which are replicated with well documented positive rechallenges (i.e. close temporal proximity between each drug exposure and an objective event that is not a manifestation or complication of the treatment indication), corroborated by robust epidemiological methods, independently detected in large postmarketing clinical trials and/or supported by cogent clinical pharmacological data, are reasonable proxy gold standards and can provide useful information on the ability of emerging quantitative methods to help identify actionable associations, although each may not reflect the full spectrum of phenomena encountered in pharmacovigilance).<sup>[69-72]</sup>

We believe that there are spontaneously reported events, largely ignored in this context, that may constitute a valuable component of reference sets used to measure data-mining performance, in that they are definitive adverse reactions that can be regarded as being pure gold or of extremely high-grade ore.

### 4. Definitive Anecdotal Adverse Drug Reactions

As sources of pure gold, we have specifically suggested four groups of spontaneously reported adverse events, for which causal or contributory attribution to the drug is either irrefutable or demonstrated with a high level of confidence.<sup>[1]</sup> The four groups are listed, with examples, in table I.<sup>[2-44]</sup> We propose that the principles illustrated define a set of gold standards that can be used to supplement those requiring clinical judgment and those defined with reproducible results in independent samples.

It should be borne in mind that the four descriptive categories and corresponding examples are not mutually exclusive or exhaustive. For example, injection-site reactions to aluminium-containing vaccines<sup>[16]</sup> have characteristics that are consistent with more than one category, since there is anatomical contiguity (category 2) and electron microprobe analysis has shown features of aluminium crystal-storing histiocytosis (category 1). Fixed drug eruptions add an additional element of physical specificity

ty that can further strengthen associations demonstrated with traditional provocative rechallenge.

In all cases, additional evidence (for example, a convincing time course) may be adduced to boost the strength or 'caratage' of an association. For many of the tests, scenarios, and procedures described in the following sections, such as photopatch testing and physiological testing of sweat and salivary gland function, positive rechallenge can also contribute, in which case blinding and the use of a placebo can further increase the evidentiary value of a single case. For example, positive rechallenge in an individual case is enhanced when the patient is unaware of the re-administration<sup>[73]</sup> or when placebo-controlled rechallenge is used.<sup>[74]</sup>

#### 4.1 Extracellular or Intracellular Tissue Deposition of the Drug or Metabolite

The first category includes adverse events in which the injury is due to either extracellular or intracellular deposition of a drug or metabolite (i.e. the pathological lesion is composed of the drug or metabolite) as demonstrated by objective physicochemical testing. The feasibility of such testing in these cases usually implies that the lesion itself is either extrinsic to body tissues or involves tissues or body fluids that are accessible for biopsy or some form of *in situ* examination. These events can be considered pure gold in the counterfactual sense,<sup>[75]</sup> in that the events could not have occurred in the absence of the drug.

It is important to recognise the distinction between a lesion caused by the compound and one in which the compound is an innocent bystander. By way of example, renal stones associated with efavirenz have been described as containing 50% metabolites of efavirenz and 50% unspecified proteins;<sup>[76]</sup> this may not be enough to exclude an innocent bystander effect, since the primary phenomenon may have been protein concretion. In well documented cases in which the calculus is composed entirely or predominantly of drug or metabolite, as has been reported for example with triamterene,<sup>[5,6]</sup> one can exclude pure confounding and infer either that an adverse drug reaction has occurred or

at least that the drug played a major contributory role by way of a drug-disease interaction.

#### 4.2 A Specific Anatomical Location or Pattern of Injury

The second category includes adverse events in which the anatomical location and/or pattern of injury is sufficiently specific to attribute the effect to the drug without the need for implicit judgment or formal investigation. The mechanism of injury can be related to physicochemical or pharmacological properties of the drug.

It is tempting to add certain intramuscular injection site reactions in this category, but caution is warranted. A distinctive example is Nicolau's syndrome (embolia cutis medicamentosa), a rare, acute, necrotic, livedoid dermatitis reported with intramuscular injection of various drugs, including bismuth, modified-release formulations of penicillin, NSAIDs and glucocorticoids.<sup>[77-80]</sup> Experimental evidence supports the theory of inadvertent intra-arterial or paravascular injection, leading to structural or functional vascular occlusion, but the relative contributions of needle injury, volume effect and/or physicochemical characteristics of the drug (for example, fat emulsion or microcrystal deposition) have not been conclusively determined. Similarly, reports of injection-site reactions with intramuscular and subcutaneous injections of drugs that are contained in bottles or cartridges containing latex plungers and diaphragms may reflect latex allergy rather than a reaction to the drug.<sup>[81,82]</sup> Nevertheless, in these cases, the event could be classified as a definitive adverse reaction to the whole formulation rather than to the drug itself.

#### 4.3 Physiological Dysfunction or Direct Tissue Damage Demonstrable by Physicochemical Testing

The third category includes adverse events that involve physiological dysfunction or tissue damage for which documentation by physicochemical testing is feasible. Drug-event associations in this category may not all be pure gold, in that some of the confirmatory test procedures may not be foolproof

and may be operator or situation dependent. However, when properly performed and interpreted they can provide a level of confidence in attribution suitable for informed decision making in pharmacovigilance. An example is photopatch testing for photo-allergy. False-negative results are problematic in this setting: a positive result could reflect a local irritant effect of the drug interacting additively with a subclinical effect of the ultraviolet light source. Nevertheless, when drug versus vehicle and irradiated versus non-irradiated controls are properly used, these tests have been reported as being diagnostic in individual cases.

#### 4.4 Infection, Either Due to the Administration of an Infective Agent as the Therapeutic Substance or to Demonstrable Contamination

The fourth category includes adverse drug reactions related to infections. These can be due either to the contaminating presence of the organism or because the product itself consists of live microbes. Confirming causality would involve proving identity between the infecting organism and the organism contained in the product and/or confirmation of matching batch-specific distribution of the contamination and the event.

## 5. Discussion

We have proposed a framework, with illustrative examples, of types of drug-event associations that can be considered to be definitive ('between-the-eyes' adverse effects) based on one or a small number of cases.<sup>[1]</sup> Preliminary data-mining analyses have already been performed for some of these associations.<sup>[6]</sup>

Given the unique character of pharmacovigilance, namely the variety of events under surveillance and the need for qualitative and quantitative probabilistic assessments at the level of single reports and causally heterogeneous case series, it is appropriate to consider how we can test the decision-making potential of pharmacovigilance systems and tools. A key question is the construction of reference sets of suspected adverse drug reactions

classified with a reasonable degree of confidence as true-positive associations. Definitive adverse reactions could contribute to that.

The adverse drug reactions we have discussed here do not represent a 'skeleton key' that will unlock the gate to a full understanding of the performance of pharmacovigilance tools and systems. They represent one subset, and a small one at that, of the 'sample space' of adverse drug reactions. One could not validate pharmacovigilance systems by extrapolating from this subset of events alone. However, the range of test scenarios and datasets should mirror the richness and variety of real-world pharmacovigilance; these types of events could form a valuable part of that range.

A fair question, along similar lines, is what is the relevance of findings from quantitative pharmacovigilance tools to reported adverse drug reactions that are definitive or highly probable based on clinical specificity? Should the application and testing of quantitative tools be limited to adverse events that are distinctive only by virtue of quantitative representation? Our response is 3-fold. If, in the real world, each and every spontaneous report from the first submission was definitive, and thus immediately recognisable, quantitative methods would offer little incremental value for prospective surveillance of these types of events. However, pharmacovigilance experience shows that complete documentation is the exception rather than the rule, and that there may be a significant time lag until the first well documented report. Secondly, a potential function of DMAs is to provide a safety net against human cognitive gaps in signal detection procedures based on manual review. Finally, the use of such events to help establish 'assay sensitivity' is related to, but distinct from, the question of how these events could be detected prospectively. In other words, the ability of quantitative tools to highlight such events may provide an additional element of reassurance in their power to detect credible phenomena. Therefore, these types of events are relevant both from the perspective of testing and for prospective surveillance.

It is also important to evaluate 'true negatives' when testing classifier technology. This is challenging, since today's 'true-negative' association could become tomorrow's 'true-positive' association in the light of additional data. We offer the following possibilities, more as a starting point for discussion rather than as foolproof solutions.

Perhaps the most basic approach is to use reported drug-event associations that are generally regarded as non-causal after extensive experience with the drug. These could suffice for real-world pharmacovigilance purposes, despite objections based on theoretical or philosophical considerations of absolute certitude. There are other possibilities. For example, collections of spontaneous reports may contain helpful data, i.e. as of the fourth quarter of 2004, there were >1500 adverse event reports listing placebo as a suspect medication. Studies of adverse non-drug events and nocebo responses are notable for the inclusion of non-serious and/or subjective events, but some of the data reported with placebo in the US FDA Adverse Event Reporting System (AERS) database are medically serious and can be associated with statistical disproportionalities.<sup>[83]</sup> These reported associations might be considered as true negatives in drug-adverse event reference sets. Although some may object that these do not truly represent spontaneous reports, spontaneous reporting system databases are plagued by numerous forms of reporting artifacts, leading to reporting that is not truly spontaneous; this is just one example in which the artifact is readily identifiable as such, and for which conclusive causality assessment can be made. In addition, some associations that were previously classified as true positives have been discounted over time (for example, 'phantom ships'),<sup>[84]</sup> and some of these may serve as true-negative reference associations. A good example is congenital anomalies reported anecdotally with Bendectin®/Debendox®<sup>1</sup> (doxylamine/dicycloverine/pyridoxine), which was thoroughly discounted over time,<sup>[85]</sup> yet has been shown to be associated with a signal of disproportionality in one preliminary data-mining exercise.<sup>[86]</sup>

Other possible true-negative reference events would be those that challenge fundamental scientific principles, i.e. the reporting of phototoxicity with potassium chloride, in which the absence of a required action spectrum would be sufficient to refute the reported association.<sup>[87]</sup>

When evaluating signal-detection methods with the aforementioned reference associations and concepts, the usual prerequisites apply: initial detection after marketing authorisation, the existence of spontaneous reports and a reasonable dictionary coding scheme to represent the reported effects. In the absence of detailed case narratives, and when the confirmed case is a literature report that is not contained in the database used for analysis, it is assumed that one or more reports of the association in the database are clinically consistent with the confirmed case. Therefore, not every example will necessarily be applicable, but collectively they can be building blocks in reference sets used to test pharmacovigilance systems. The associated concepts can be used to identify additional events worthy of consideration. We also hope that this analysis will stimulate discussions of causality assessment of individual reports and optimising clinical cognition in signal detection and evaluation.

## 6. Conclusions

In this article, we have established the notion that isolated reports can be definitive in constructing gold standards in pharmacovigilance. We stress that we are not proposing a new classification of adverse drug reactions; we have merely described categories of reaction that can be regarded as definitive when described anecdotally. Our list of categories is probably not exhaustive. In addition, although we have identified adverse effects that would not need confirmation in formal studies, studies with independent datasets would still be necessary to quantify the risk.

What we have not established here are best practices for designing and interpreting validation and testing using such gold standards. Our analysis does

**1** The use of trade names is for product identification purposes only and does not imply endorsement.



not obviate the need for continuing discussions about broader issues in the testing of emerging pharmacovigilance technologies, such as the inherent limitations of all forms of disproportionality analysis, the relevance of tests in isolation (the analogue of so-called 'test research') versus the study of the incremental contribution of the new tools to existing methods (so-called 'diagnostic research')<sup>[88]</sup> and the optimal deployment of emerging technologies within comprehensive signal detection programmes based on multiple methods and data streams.

Perhaps more fundamentally related to the issue of gold standards is the nature of a 'signal' itself. Some formulations of signal detection performance are based on the availability of associations that are provable without any residual uncertainty. We think this may be a limited and unproductive approach that misses the value of 'situational awareness' in real-world pharmacovigilance.<sup>[66]</sup>

Spontaneous anecdotal reports have limitations, but we believe that they can also be rich sources of gold nuggets, which can be mined for possible inclusion in reference sets for testing any method used to support the signal detection process. This is in the spirit of others who have discussed the tendency to undervalue observations from case reports.<sup>[89]</sup> Because the types of events we describe are diverse but relatively uncommon, pure gold and high-grade ores do not substitute for reference events traditionally used to test signal-detection tools. By themselves they would constitute a quantitatively inadequate and biased reference set. Rather, we believe that they can supplement and enrich the usual approaches.

## Acknowledgements

We are grateful for the helpful comments of David Madigan, Robin Ferner, Paul Glasziou, Yoon Loke and Jan Vandenbroucke, who reviewed draft versions of the manuscript, and Andrew Bate, François Girardin and Sheila Weiss Smith, who refereed our original paper in the *British Medical Journal*.

Dr Hauben is an employee of Pfizer Inc. and owns stock and stock options in Pfizer and other pharmaceutical companies. The authors have no conflicts of interest that are directly

relevant to the content of this manuscript. No funding was provided for the preparation of this review.

## References

1. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. *BMJ* 2006; 333: 1267-9
2. Biscckford CL, Spencer AP. Biliary sludge and hyperbilirubinaemia associated with ceftriaxone in an adult: case report and review of the literature. *Pharmacotherapy* 2005; 25 (10): 1389-95
3. Tokumine F, Sunagawa T, Shiohira Y, et al. Drug-associated cholelithiasis: a case of sulindac stone formation and the incorporation of sulindac metabolites into the gallstones. *Am J Gastroenterol* 1999; 94 (8): 2285-8
4. Eda A, Yanaka I, Tamada K, et al. Sulindac-associated choledocholithiasis. *Am J Gastroenterol* 2001; 96 (7): 2283-5
5. Daudon M, Jungers P. Drug-induced renal calculi: epidemiology, prevention, and management. *Drugs* 2004; 64 (3): 245-75
6. Hauben M, Reich L, Gerrits C. Comparative performance or proportional reporting ratios (PRR) and multi-item gamma-Poisson shrinker (MGPS) for the identification of crystalluria and urinary tract calculi caused by drugs. *Pharmacoepidemiol Drug Saf* 2005; S014: 7
7. Messmer E, Font RL, Sheldon G, et al. Pigmented conjunctival cysts following tetracycline/minocycline therapy. *Ophthalmology* 1983; 90: 1462-8
8. Morrison VL, Kikkawa DO, Herndier BG, et al. Tetracycline induced green conjunctival deposits. *Br J Ophthalmol* 2005; 89: 1372-3
9. Bandla HPR, Davis SH, Hopkins NE. Lipoid pneumonia: a silent complication of mineral oil aspiration. *Pediatrics* 1999; 103 (2): E19
10. Taylor JR, Streetman DS, Castle SS. Medication bezoars: a literature review and a report of a case. *Ann Pharmacother* 1998; 32 (9): 940-6
11. Champman AH, el Hasani S. Colon ischaemia secondary to barolith obstruction. *Br J Radiol* 1998; 71 (849): 983-4
12. Eiferman RA, Snyder JP, Nordquist RE. Ciprofloxacin microprecipitates and macroprecipitates in the human corneal epithelium. *J Cataract Refract Surg* 2001; 27: 1701-2
13. Lopez JD, del Castillo JMB, Lopez CD, et al. Confocal microscopy in ocular chrysiasis. *Cornea* 2003; 22 (6): 573-5
14. Nadin F, Haddad W, Adib J. The differential diagnosis of crystals in the retina. *Int Ophthalmol* 2001; 24 (3): 113-21
15. Goralczyk R, Barker FM, Buser S, et al. Dose dependency of canthaxanthin crystals in monkey retina and spatial distribution of its metabolites. *Invest Ophthalmol Vis Sci* 2000; 41 (6): 1513-22
16. Culora GA, Ramsay AD, Theaker JM. Aluminium and injection site reactions. *J Clin Pathol* 1996; 49 (10): 844-7
17. Sukpanichnant S, Hargrove NS, Kachintorn U, et al. Clofazimine-induced crystal-storing histiocytosis producing chronic abdominal pain in a leprosy patient. *Am J Surg Pathol* 2000; 24 (1): 129-35
18. Zanetta G, Maurice-Estepa L, Mousson C, et al. Foscamet-induced crystalline glomerulonephritis with nephritic syndrome and acute renal failure after kidney transplantation. *Transplantation* 1999; 67 (10): 1376-8
19. Rollins SD, Craig JP. Gold-associated lymphadenopathy in a patient with rheumatoid arthritis. Histologic and scanning electron microscopic features. *Arch Pathol Lab Med* 1991; 115 (2): 175-7

20. Hendricks A. Yellow lunulae with fluorescence after tetracycline therapy. *Arch Dermatol* 1980; 116: 438-40
21. Dixit VB, Chaudhary SD, Jain VK. Clofazimine induced nail changes. *Indian J Lepr* 1989; 61: 476-8
22. Adams PC, Holt DW, Storey GCA, et al. Amiodarone and its desethyl metabolite: tissue distribution and morphologic changes during long-term therapy. *Circulation* 1985; 72 (5): 1064-75
23. O'Neill JL, Remington TI. Drug-induced esophageal injuries and dysphagia. *Ann Pharmacother* 2003; 37 (11): 1675-84
24. Adami NP, de Gutierrez MG, da Fonseca SM, et al. Risk management of extravasation of cytostatic drugs at the Adult Chemotherapy Outpatient Clinic of a university hospital. *J Clin Nurs* 2005; 14 (7): 876-82
25. van der Leede H, Jorens PG, Parizel P, et al. Inadvertent intrathecal use of ionic contrast agent leading to nervous system toxicity. *Eur Radiol* 2002; 12 Suppl. 3: S86-93
26. Alcaraz A, Rey C, Conch A, et al. Intrathecal vincristine and fatal myeloecephalopathy despite cerebrospinal fluid perfusion. *J Toxicol Clin Toxicol* 2002; 40 (5): 557-61
27. Manson AJ, Hanagasi H, Turner K, et al. Intravenous apomorphine therapy in Parkinson's disease: clinical and pharmacokinetic observations. *Brain* 2001; 124 (Pt 2): 331-40
28. Sapir S, Bimstein E. Cholinergic dentate gel induced oral lesion: report of case. *J Clin Pediatr Dent* 2000; 24 (2): 103-6
29. Brazier WJ, Dhariwal DK, Patton DW, et al. Ecstasy related periodontitis and mucosal ulceration: a case report. *Br Dent J* 2003; 194 (4): 197-9
30. Seyer BA, Grist W, Muller S. Aggressive destructive midfacial lesion from cocaine abuse. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2002; 94 (4): 465-70
31. Yamasaki K, Yamasaki A, Tosaki M, et al. Tissue distribution of Thorotrast and role of internal irradiation in carcinogenesis. *Oncol Rep* 2004; 12 (4): 733-8
32. Ben-Zeev B, Watemberg N, Augarten A, et al. Oligohydrosis and hyperthermia: a pilot study of a novel topiramate adverse event. *J Child Neurol* 2003; 18 (4): 254-7
33. Shimizu T, Yamashita Y, Satoi M, et al. Heat stroke-like episode in a child caused by zonisamide. *Brain Dev* 1997; 19: 366-8
34. Okumura A, Hayakawa F, Kuno K, et al. Oligohydrosis caused by zonisamide. *No To Hattatsu* 1996; 28(1): 44-7
35. Lee A, Joo H, Chey W, et al. Photopatch testing in seven cases of photosensitive drug eruptions. *Ann Pharmacother* 2002; 35: 1584-7
36. Martin-Lazaro J, Bujan JG, Arrondo AP, et al. Is photopatch testing useful in the investigation of photosensitivity due to flutamide? *Contact Dermatitis* 2004; 50 (5): 325-6
37. Jeanmougin M, Manciet JR, De Prost Y, et al. Fenofibrate photoallergy. *Ann Dermatol Venerol* 1993; 120 (8): 549-54
38. Schiffman SS, Zervakis J, Westall HL, et al. Effects of antimicrobial and anti-inflammatory medications on the sense of taste. *Physiol Behav* 2000; 69: 413-24
39. Teare JP, Spedding C, Whitehead MW, et al. Omeprazole and dry mouth. *Scand J Gastroenterol* 1995; 30 (3): 216-8
40. Trevenzoli M, Cattelan AM, Marino F, et al. Sepsis and granulomatous hepatitis after bacillus Calmette-Guérin intravesical installation. *J Infect* 2004; 48 (4): 363-7
41. Land MH, Rouster-Stevens K, Woods CR, et al. Lactobacillus sepsis associated with probiotic therapy. *Pediatrics* 2005; 115 (1): 178-81
42. Kashiwagi Y, Kawashima H, Takekuma K, et al. Detection of mumps virus genome directly from clinical samples and a simple method for genetic differentiation of the Hoshino vaccine strain from wild strains of mumps virus. *J Med Virol* 1997; 52: 195-9
43. Hauben M, Reich L. Endotoxin-like reactions with intravenous gentamicin: results from pharmacovigilance tools under investigation. *Infect Control Hosp Epidemiol* 2005; 26 (4): 391-4
44. Bennett SN, McNeil MM, Bland LA, et al. Postoperative infections traced to contamination of an intravenous anesthetic, propofol. *N Engl J Med* 1995; 333 (3): 147-54
45. Koh Y, Shu CL. A new algorithm to identify the causality of ADRs. *Drug Saf* 2005; 28 (12): 1159-61
46. Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. *Lancet* 2000; 356 (9237): 1255-9
47. Valenstein PN. Evaluating diagnostic tests with imperfect standards. *Am J Clin Pathol* 1990; 93: 252-8
48. Obuchowski NA. Special topics III: bias. *Radiology* 2003; 229: 617-21
49. Streiner DL. Diagnosing tests: using and misusing diagnostic and screening tests. *J Pers Assess* 2003; 81 (3): 209-19
50. Coggon D, Martyn C, Palmer KT, et al. Assessing case definitions in the absence of a diagnostic gold standard. *Int J Epidemiol* 2005; 34: 949-52
51. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol* 1993; 137: 1251-8
52. Bossuyt PM, Reitsma JB, Bruns DE, et al. Standards for reporting of diagnostic accuracy. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem* 2003; 49 (1): 7-18
53. Pfeiffer RM, Castle PE. With or without a gold standard. *Epidemiology* 2005; 16 (5): 595-7
54. Greenland S. Multiple-bias modeling for analysis of observational data. *J R Stat Soc* 2005; 168: 267-308
55. Phillips CV. Quantifying and reporting uncertainty from systematic errors. *Epidemiology* 2003; 14: 459-66
56. Aronson JK, Ferner RE. Clarification of terminology in drug safety. *Drug Saf* 2005; 28 (10): 851-70
57. Hauben M, Reich L, Gabbay F. Extension of points on clarifying terminology in drug safety. *Drug Saf* 2006; 29 (3): 273-5
58. Buehler JW, Hopkins RS, Overhage JM, et al. Framework for evaluating public health surveillance systems for early detection of outbreaks. *MMWR Recomm Rep* 2004; 53 (RR-5): 1-11
59. Aronson JK. Unity from diversity: the evidential use of anecdotal reports of adverse drug reactions and interactions. *J Eval Clin Pract* 2005; 11 (2): 195-208
60. Hauben M, Madigan D, Gerrits C, et al. The role of data mining in pharmacovigilance. *Exp Opin Drug Saf* 2005; 4 (5): 929-48
61. Hauben M, Patadia V, Gerrits C, et al. Data mining in pharmacovigilance: the need for a balanced perspective. *Drug Saf* 2005; 28 (10): 835-42
62. Banks D, Woo EJ, Burren DR, et al. Comparing data mining methods in the VAERS database. *Pharmacoepidemiol Drug Saf* 2005; 14 (9): 601-9
63. Lilienfeld DE. A challenge to data miners. *Pharmacoepidemiol Drug Saf* 2005; 13 (12): 881-4
64. Hauben M, Bate A. Data mining in drug safety. In: Aronson JK, editor. *Side effects of drugs annual 29*. Amsterdam: Elsevier Science Ltd, 2007: xxxiii-xlvi
65. Hauben M, Reich L. Response to Levine et al. *Br J Clin Pharmacol* 2006; 61 (1): 115-7

66. Fricker RD, Rolka H. Protecting against biological terrorism: statistical issues in electronic biosurveillance. *Chance* 2007; 19 (4): 4-13
67. Tramèr MR, Moore RA, Reynolds DJ, et al. Quantitative estimation of rare adverse events which follow a biological progression: a new model applied to chronic NSAID use. *Pain* 2000; 85 (1-2): 169-82
68. Wald NJ, Morris JK. Teleoanalysis: combining data from different types of study. *BMJ* 2003; 327 (7415): 616-8
69. Hauben M. Trimethoprim-induced hyperkalemia: lessons in data mining. *Br J Clin Pharmacol* 2004; 58: 338-9
70. Hauben M, Reich L. Drug-induced pancreatitis: lessons in data mining. *Br J Clin Pharmacol* 2004; 58 (5): 560-2
71. Bate A, Lindquist M, Orre R, et al. Data-mining analyses of pharmacovigilance signals in relation to relevant comparison drugs. *Eur J Clin Pharmacol* 2002; 58 (7): 483-9
72. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-normal expected combinations of drugs and events in the FDA spontaneous reports database. *Drug Saf* 2002; 25 (6): 381-92
73. Korantzopoulos P, Pappa E, Karanikis P, et al. Acute low-back pain during intravenous administration of amiodarone: a report of two cases. *Int J Cardiol* 2005; 98 (2): 355-7
74. Smyth J. Determining optimal therapy-randomized trials in individual patients. *N Engl J Med* 1986; 315: 767-8
75. Hofler M. Analytical perspective. The Bradford Hill considerations on causality: a counterfactual perspective. *Emerg Themes Epidemiol* 2005; 2: 11
76. Wirth GJ, Teuscher J, Graf JD, et al. Efavirenz-induced urolithiasis. *Urol Res* 2006; 34: 288-9
77. McGee AM, Davison PM. Skin necrosis following injection of non-steroidal anti-inflammatory drug. *Br J Anaesthesia* 2002; 88: 139-40
78. Cherasse A, Kahn M, Mistrih R, et al. Nicolau's syndrome after local glucocorticoid injection. *Joint Bone Spine* 2003; 70 (5): 390-2
79. Kohler LD, Schwedler S, Worret WI. Embolia cutis medicamentosa. *Int J Dermatol* 1997; 36 (3): 197
80. Albrecht KH, Littman K, Richter HJ, et al. Animal experiment studies of embolia cutis medicamentosa. *Langenbecks Arch Chir* 1984; 362 (1): 17-23
81. Shojaei AR, Haas DA. Local anesthetic cartridges and latex allergy: a literature review. *J Can Dent Assoc* 2002; 68 (10): 622-6
82. Towse A, O'Brien M, Twarog FJ, et al. Local reaction secondary to insulin injection: a potential role for latex antigens in insulin vials and syringes. *Diabetes Care* 1995; 18 (8): 1195-7
83. Hauben M. Data mining in pharmacovigilance: integrating statistical and computational approaches. *Drug Safety Research Unit 3rd Biennial Conference on Signal Detection and Evaluation*; 2005 Sep 28-29; London
84. Stricker B. Pharmacovigilance: a case of phantom ships and Russian roulette. *Ned Tijdschr Geneesk* 2002; 146 (27): 1258-61
85. Mazzotta P, Magee LA. A risk-benefit assessment of pharmacological and nonpharmacological treatments for nausea and vomiting of pregnancy. *Drugs* 2000; 59 (4): 781-800
86. Hauben M, Reich L, van Puijenbroek, et al. Data mining in pharmacovigilance: lessons from phantom ships. *Eur J Clin Pharmacol* 2006; 62: 967-70
87. Moore D. Drug-induced cutaneous photosensitivity: incidence, mechanism, prevention and management. *Drug Saf* 2002; 25 (5): 345-72
88. Moons KGM, Biesheuvel CJ, Grobbee DE. Test research versus diagnostic research. *Clin Chem* 2004; 50 (3): 473-6
89. Vandembroucke JP. In defense of case reports and case series. *Ann Intern Med* 2001; 134 (4): 330-4

---

Correspondence: Dr *Manfred Hauben*, Risk Management Strategy, Pfizer Inc., 235 East 42nd Street, New York, NY 10017, USA.

E-mail: Manfred.hauben@pfizer.com