

# 1

## Introduction

### A. Lawrence Gould

*Merck Research Laboratories, 770 Sumneytown Pike, West Point, PA 19486, USA*

#### 1.1 Introduction

Many stakeholders have an interest in how pharmaceutical products are developed. These include the medical profession, regulators, legislators, the pharmaceutical industry, and, of course, the public who ultimately will use the products. The expectations of these stakeholders have become more demanding over time, especially with regard to product safety. The public perception of the safety of pharmaceutical products often is driven by publicity about the occurrence of adverse events among patients using the products that has on occasion led to withdrawal of the products from the market [1]. This circumstance usually pertains to products that have reached the marketplace and have had sufficient exposure among patients for rare and potentially serious harmful events to occur frequently enough to cause concern. However, the development of products can be suspended or terminated before they ever reach the market because of toxicities discovered during development [2–6]. These situations may or may not be made the object of intense public scrutiny, but they are important because failed products can have consumed possibly considerable resources that might have been allocated more productively to the development of products more likely to succeed by virtue of being less toxic or more beneficial.

Any biologically active pharmaceutical product potentially can harm as well as benefit its users. This can happen because a drug or biological agent has multiple mechanisms of action besides those involved in the therapeutic target, or idiosyncratically, possibly because of an immune response. It also can happen because of how the body reacts to non-pharmaceutical products, especially indwelling medical devices such as cardiovascular stents or artificial joints. Understanding how these potential harms can manifest themselves and at what stages of product development the potential for harm can be identified is critical to the development

---

*Statistical Methods for Evaluating Safety in Medical Product Development*, First Edition.

Edited by A. Lawrence Gould.

© 2015 John Wiley & Sons, Ltd. Published 2015 by John Wiley & Sons, Ltd.



## 2 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

of products that provide real benefits to patients. Many potential products that enter development fail because of unanticipated safety issues. Some of these occur early in development, but some occur very late in development. It is important to be able to predict the likelihood of harm from potential products as early as possible in the development process, and certainly before they reach the marketplace and present unnecessary risks to large numbers of patients.

### 1.2 Background and context

The safety of drugs, vaccines, and medical devices has become the Pole Star of product development. Discovery and development of drugs and other pharmaceutical products takes a long time, costs a lot of money, and has a low probability of success [7]. Failures can occur often during the development process, especially for novel drugs [2, 3]. Product withdrawals also can occur after products have been approved for marketing although, adverse publicity notwithstanding, these are relatively rare. Of the 740 new molecular entities (NMEs) approved by the Food and Drug Administration (FDA) in the USA between 1980 and 2009, 118 were withdrawn from the market. Most of these withdrawals were for reasons other than safety. Only 26 NMEs (3.5% of the approvals during this period) were withdrawn for safety reasons [8].

Safety issues arising consequent to chronic treatment do not always appear evident during drug development, either by preclinical assays or in the clinical phase of development. At least for cardiovascular events there is a need for understanding of fundamental mechanisms of cardiovascular liability that provide a way to detect potential toxicities during development [9]. The possibility of using biomarkers as leading indicators of potential safety issues has become a subject of discussion in the recent literature [10, 11]. There also has emerged in recent years an increasing interest in the application of methods for preclinical safety pharmacology and computational toxicology [12–16].

There is an increasing appreciation and availability of sophisticated means for making measurements early in the drug development process to identify potential safety issues that may emerge later on. There also is a need for means to provide more realistic assessments of risks of adverse events than are provided by clinical trials that do not, ordinarily cannot, include patients across the spectrum of potential susceptibility to adverse events [17–19].

Advances in the sophistication of measurement and interpretation of data make it appropriate to consider how recent developments in statistical methods for modeling, design, and analysis can contribute to progress in drug development, especially with regard to evaluating safety. Many books, and many more articles, describe conventional strategies for evaluating the safety of pharmaceutical products at various stages of development. Balakrishnan *et al.* [20] provide an exhaustive collection (86 chapters) of statistical methods but without a focus on safety. Chow and Liu [21] focus on the design of clinical trials, but do not address safety in depth or describe the implementation of novel methods for dealing with new types of complex data. Everitt and Palmer [22] provide an exhaustive collection of statistical essays intended to give medical researchers and clinicians readable accounts statistical concepts as they apply in various areas of medical research, especially in various therapeutic areas. Gad [23] focuses primarily on non-clinical pharmacology and toxicology studies needed to support product development with some attention to safety assessment in humans during and after the clinical development process, but does not appear to be directed toward statistical methods that can be applied, except possibly for conventional methods; there do not appear to be



any references to the statistical literature past 1994. Lachin [24] describes standard tools and more recent likelihood-based theories for assessing risks and relative risks in clinical investigations, especially two-group comparisons, sample size considerations, stratified-adjusted analyses, case-control and matched studies, and logistic regression. Moyé [25] covers a number of topics, but addresses safety fairly briefly from a monitoring point of view. Proschan *et al.* [26] address the theoretical and practical aspects of monitoring clinical trials, primarily with the aim of assessing efficacy, but also with recommendations for monitoring safety in ongoing trials. The book edited by Rao *et al.* [27] covers a substantial range of methods for addressing various aspects of the design and analysis of clinical trials, including early phase trials and post-marketing trials, but does not address safety as such. Senn [28] identifies and addresses various issues, including (briefly) safety.

### 1.3 A fundamental principle for understanding safety evaluation

The evaluation of efficacy differs fundamentally from the evaluation of safety of medical products, that is, drugs, vaccines, and medical devices. To simplify the presentation in what follows, the term “drug” or “therapy” generally be used; however, statements using these terms generally will apply to any medical product.

Efficacy is at least conceptually easy to evaluate because the criteria for assessing efficacy in a trial need to be specified explicitly at the outset. A trial is designed with the expectation that one or more specific null hypotheses of no difference between the effect of the test therapy and a control will be rejected on the basis of the observations made during the trial. An antidiabetic drug may be assessed in terms of change in HbA1c over a defined period of time, an antiarrhythmic drug may be assessed in terms of survival, an antidepressant may be assessed in terms of change in Hamilton Depression Rating Score after a few months of treatment, and so on. If there is more than one hypothesis to be tested, some adjustment for the fact that multiple tests are performed is made in the statistical analysis so that the probability of concluding that a test therapy is efficacious when it is not can be controlled at an acceptable level. What constitutes efficacy and what the expectations are at the outset are known. That is how the sample size for the trial is determined. This is the same whether the aim of the trial is to prove that a new therapy is superior to a control or, if not, that it is not materially inferior to the control.

Safety is different. Although a few hypotheses about specific safety issues can be identified at the outset of a trial, and are treated in the statistical analysis similarly (but not identically) to hypotheses about efficacy, most safety issues are not identified at the outset of the trial. Consequently, the basis for determining that a test therapy is or is not acceptably “safe” generally cannot be identified before undertaking the trial. The inference about safety rests on interpretation of the observations. This can be problematic for at least two reasons. Firstly, it amounts to using the same observations to generate and to test hypotheses, which violates a basic scientific principle [29]. Secondly, attempts to adjust for the multiplicity of tests that are carried out for the often substantial number of adverse events that emerge during a trial using the same approaches that would apply for evaluating efficacy can decrease the sensitivity of any comparison so much that no difference in toxicity risk can be detected. However, not adjusting for multiplicity means that the chance of finding a material difference between the test and control therapies becomes appreciable even when the therapies pose the same risks.

#### 4 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

How is one to interpret a “significant” increase in cardiac arrhythmias on a test therapy when perhaps 50 different adverse events that were not identified at the outset emerge during the trial? Is this a real effect, or is it a statistical artifact due to the fact that 50 tests were carried out? Clearly, a test of the null hypothesis that there is no additional risk of arrhythmia cannot by itself *confirm* an elevated risk of arrhythmia, but the way the findings from the trial usually are presented to the medical world at large invites a statistically significant finding to be (mis)interpreted as demonstrating at least association if not outright causality [30–32].

### 1.4 Stages of safety evaluation in drug development

Consideration of the potential toxicity of a potential drug or vaccine occurs at every stage of development:

1. Preclinical (efficacy, toxicity, pharmacokinetics, and epidemiology)
  - a. *In silico* (computational toxicology, quantitative structure–activity relationship, and chemometrics)
  - b. *In vitro* studies
  - c. *In vivo* studies
2. Phase 1 (first in humans, healthy populations–toxicity and pharmacokinetics/pharmacodynamics)
3. Phase 2 (toxicity and efficacy)
  - a. Proof of concept studies
  - b. Dose ranging studies
4. Phase 3 (toxicity and efficacy)
  - a. Randomized controlled trials
  - b. Confirmation of hypotheses generated in Phase 2
5. Post-approval
  - a. Phase 4 studies
  - b. Post-marketing safety surveillance
  - c. Pharmacoepidemiology.

At every stage, the aims are to identify and characterize potential safety problems, understand the risk–benefit balance, and (especially at later stages) plan for risk management and mitigation.

While toxicity may be manifested in a variety of ways, certain key potential safety issues pervade the developmental process. These include cardiotoxicity, especially alterations of the electrical activity of the heart as manifested on electrocardiograms (ECGs), hepatotoxicity, nephrotoxicity, and bone marrow toxicity. Drugs in particular can be metabolized in different ways, and can have effects that depend on, that is, interact with, other co-administered drugs, the patient’s disease state, and the general environment of the patient’s life. Many adverse

events can occur, but serious adverse events are potentially of most concern, particularly death, life-threatening events (e.g., bone marrow suppression), events leading to hospitalization, events leading to significant, persistent, or permanent disability, and congenital anomalies and other birth defects.

## 1.5 National medical product safety monitoring strategy

The National Medical Product Safety Monitoring Strategy that is part of the FDA Sentinel Initiative [33] aims to provide an integrated approach to monitoring pharmaceutical product safety throughout the entire life cycle of a product. The strategy seeks to combine an understanding of the underlying disease states with new methods of signal detection, data mining, and analysis. The Organizational Medical Outcomes Partnership, established under the Foundation of the National Institutes of Health, unites regulatory, industrial, and academic contributors in the development and implementation of tools for carrying out the monitoring strategy on marketed products using medical information accumulated in a variety of databases.

The overall strategy is implemented through an interdisciplinary team approach including geneticists, cell biologists, clinical pharmacologists, statisticians, epidemiologists, and informatics experts. The teams are charged with generating and confirming hypotheses about causal factors of safety problems among product users.

## 1.6 Adverse events vs adverse drug reactions, and an overall view of safety evaluation

Adverse events are not the same as adverse drug reactions. The ICH Guidelines (E6) define an *adverse event* as “any untoward medical occurrence in a patient or subject administered in a pharmaceutical product” [34]. Untoward medical occurrences include any unfavorable and unintended sign, symptom, or disease temporally associated with the use of the product. It is only the occurrences that matter, without a judgment about causality. The same guidelines define *adverse drug reactions* as either all noxious and unintended responses at least possibly causally related to a new medicinal product or usage, or a noxious and unintended response that occurs to patients on a marketed medicinal product at doses normally used in man. The element of possible causality is the key difference between these two concepts. There is, in addition, an extensive set of safety guidelines [35].

Risks associated with medicinal products arise from a number of sources, diagrammed in Figure 1.1. Some risks are known consequences of administration of medicinal products due to the disease treated or to the mechanism of action. Of these, some are unavoidable consequences that can lead to injury or death, and are part of the cost of the potential benefits of the product. Others are avoidable by appropriate choice of dosage or identification of patients who should or should not receive the product. Avoidable adverse events are in principle at least preventable, so that a key focus of safety evaluation is determining how characteristics of patients and dosage strategies are related to the risk of preventable adverse events. However, adverse events can occur for other than predictable reasons, especially when usage of the product is extended to wider patient populations than were studied during product development, or to uses not included in the originally approved set of indications.

6 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

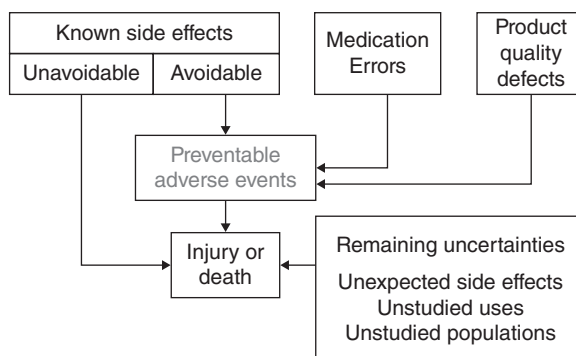


Figure 1.1 Sources of risk from medicinal products.

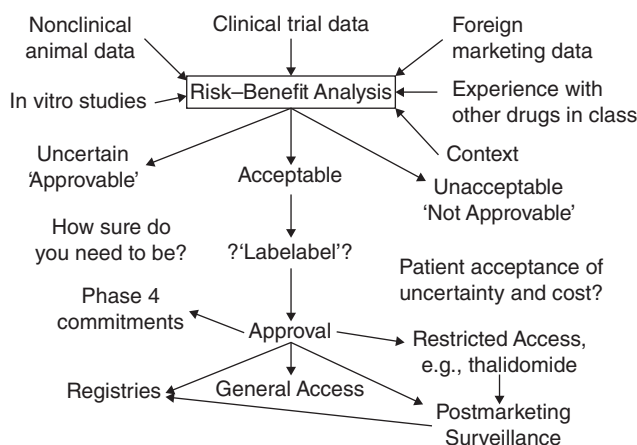
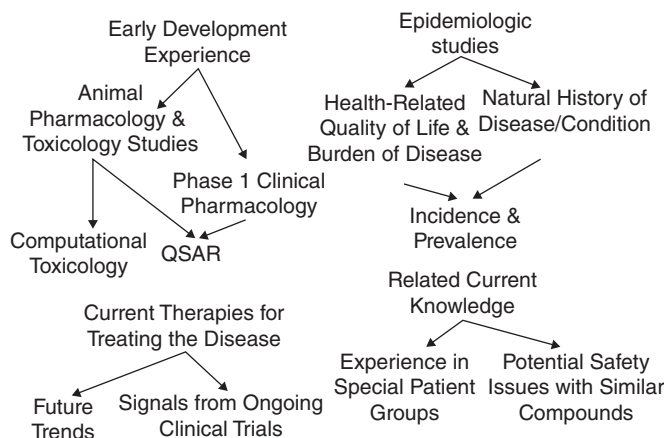


Figure 1.2 Overview of the components of the risk-benefit evaluation process in a regulatory context.

The evaluation of medicinal product safety occurs in the larger context of risk management and risk-benefit assessment. Figure 1.2 provides an outline of the context. The evaluation of risk and benefit incorporates information from clinical trials, but also information from non-clinical animal and *in vitro* studies that may provide insights into mechanisms of action, experience from other drugs in the same class, and the context for the use of the drug. Risk-benefit analyses lead to recommendations for further actions. If the risks are unacceptable, then the drug will be deemed “not approvable.” If the risks are acceptable then the approval process can proceed to the next steps. If there is uncertainty about the risks, then further evaluation or trials may be needed. Once a drug has reached the point of possible approval, further decisions are in order, for example, whether the drug should be approved for general access or restricted access because there is a subpopulation for which the risk is unacceptable. In addition, approval may be granted conditional on further studies to assure that important, but rare, risks are not overlooked. In some cases, registries may be set up to follow specific subgroups of patients, for example, women who become pregnant. For most



**Figure 1.3** Elements of strategies for safety evaluation.

drugs, post-marketing surveillance of at least spontaneous reports will be required as part of periodic safety updates.

As the overview in Figure 1.2 makes clear, the evaluation of safety for any medicinal product needs to be considered in a fairly wide context and, therefore, must be the consequence of a well-thought-out strategy. Figure 1.3 illustrates some of the considerations that drive strategies for evaluating safety.

## 1.7 A brief historical perspective on safety evaluation

The discovery, testing, and utilization of medicines to treat human and animal ailments is as ancient as human history. Every human culture has its pharmacopeia based on extensive, if haphazard, trial and error [36–44]. Not all medications are derived from herbs. Some are derived from animals [40], some from marine fauna [44], and some from minerals and metals [45, 46]. Many ancient medicines whose effectiveness rests on long traditions of observation present potential health issues because of their pharmacologic effects, so that their use often traditionally has been restricted to physicians or trained healers [47]. Most traditional medicines have not been assessed by the standards employed for more conventional modern therapies so that even though they may be effective for specific purposes, they also may be toxic or may interact with conventional therapies in ways injurious to a patient [47–55].

The need to assure the safety of medicines and medical devices, a key objective of modern product development, has been recognized for centuries [56, 57]. For example, of the components that traditional Chinese medicine principles identify for a medicinal compound, one (the “adjuvant”) is specifically intended to neutralize any side effects of the effective moieties (the “monarch” and the “minister”) and another (the “guide”) is intended essentially to enhance their bioavailability [39]. However, effective attempts to provide this assurance by government regulation of these products, at least in the West, date back, with apparently one exception, only to the late eighteenth and nineteenth centuries [58]. The exception, which addresses quality control of the preparation process rather than potential toxicity, is a pair of compounds formulated in the second century BCE, mithridatium and theriac, collectively

## 8 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

known as “Venetian treacle,” that led to a series of statutes in England dating from 1540 through 1799 for the regulation of the production of pharmaceuticals [47].

Although every national government includes a ministry of health (possibly with a different name) whose responsibilities include some form of evaluation of medicinal products, oversight of drug product safety is (or, at least until recently, has been) essentially non-existent in many parts of the Third World [59].

In the United States, the Biologics Control Act of 1902 authorized the regulation of the sale of biologic agents (viruses, sera, toxins, etc.); this Act required licensing of manufacturers and manufacturing establishments, established standards for safety, purity, and potency of biologics, and gave the federal government inspection authority. The Pure Food and Drug Act of 1906 culminated over 40 years of effort and various Acts of Congress to produce effective regulatory oversight of food and medicines; this Act prohibited interstate commerce in misbranded and adulterated foods, drinks, and drugs [60]. The 1906 legislation had a number of legal and regulatory shortcomings and was superseded by the Federal Food, Drug, and Cosmetic Act of 1938 after five years of legislative struggle, largely in response to a scandal resulting from the deaths of 107 children due to a lethal diluent used in a preparation of sulfanilamide elixir marketed by the Massengill Corporation [60, 61]. The new legislation required that new drugs be shown to be safe before marketing, extended regulatory control to cosmetics and therapeutic devices, and added the remedy of court injunctions to previous penalties, among other provisions. The legislation was amended several times in subsequent years. The most significant changes were implemented in 1962 with the Kefauver–Harris Drug Amendments that required drug manufacturers to prove the effectiveness of their products to the FDA before marketing. The Medical Device Amendments were passed in 1976, to assure safety and effectiveness of medical devices and diagnostic products.

The Council for International Organizations of Medical Sciences (CIOMS) was established jointly by WHO and UNESCO in 1949 as an umbrella organization for facilitating and promoting international biomedical science activities, especially those requiring the participation of international associations and national institutions. CIOMS has initiated and coordinated various long-term programs pertaining to drug development and use. CIOMS working groups have covered a broad range of drug safety topics, including consensus guidelines for reporting adverse drug reactions, drug safety update summaries, development safety update reports, core clinical safety information on drugs, terminology of adverse drug reactions, standardized MedDRA queries, and pharmacogenetics [62].

### 1.8 International conference on harmonization

The importance of independent evaluation of medicinal products before release to the public was reached at different times in different regions. Laws, regulations, and guidelines for reporting and evaluating data on safety, quality, and efficacy of new medicinal products increased rapidly in the 1960s and 1970s for most countries. Although the medicinal products industry was becoming more international, variations in technical and regulatory requirements among countries made it necessary for producers to duplicate time-consuming and expensive test procedures in order to market new products internationally. Concerns over rising costs of healthcare, escalation of the cost of research and development (R&D), and the need to meet the public expectation of rapid availability of safe and efficacious new treatments drove an urgent need to rationalize and harmonize regulatory requirements.



The European Community (now the European Union) pioneered the harmonization of regulatory requirements in the 1980s, concurrently with the evolution of a single market for medical products. Discussions initiated at the WHO Conference of Drug Regulatory Authorities in 1989 led to the establishment of a joint regulatory-industry initiative involving representatives of regulatory agencies and industry associations from Europe, Japan, and the United States that became the International Conference on Harmonisation (ICH) in 1990 [63]. The ICH steering committee decided to focus on harmonization of requirements for establishing the safety, quality, and efficacy of new medical products, leading to the ICH Harmonised Tripartite Guidelines. The effort to implement the guidelines and recommendations in the initial ICH regions and to extend the benefits to other regions has continued in the decades since the establishment of the ICH.

## 1.9 ICH guidelines

The safety guidelines provide detailed descriptions of the kinds of information that need to be obtained to evaluate various aspects of safety, as indicated in Table 1.1 [64, 65]. They often provide specific recommendations for specific kinds of experiments or trials, including recommendations about how doses should be selected. However, in contrast to considerable discussion of design and analysis considerations for evaluating efficacy, no details are provided about appropriate statistical designs as such or analytic methods, although Guideline S5 provides valuable insights:

“Significance” tests (inferential statistics) can be used only as a support for the interpretation of results. The interpretation itself must be based on biological plausibility. It is unwise to assume that a difference from control values is not biologically relevant simply because it is not “statistically significant”. To a lesser extent it can be unwise to assume that a “statistically significant” difference must be biologically relevant. ... Confidence intervals for relevant quantities can indicate the likely size of the effect.

**Table 1.1** ICH safety guidelines.

Guideline	Addresses	Guideline	Addresses
S1A, S1B, S1C	Carcinogenicity	S6	Biotechnology-derived pharmaceuticals
S2A, S2B	Genotoxicity	S7A	Safety pharmacology studies for human pharmaceuticals
S3A	Toxicokinetics	S7B	QT interval prolongation for human pharmaceuticals
S3B	Repeated dose tissue distribution	S8	Immunotoxicity studies
S4	Chronic toxicity testing in animals	S9	Non-clinical evaluation for anticancer pharmaceuticals
S5	Detection of reproductive toxicity	M3	Non-clinical safety studies to support human clinical trials

## 10 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

The efficacy guidelines are more extensive, and consider many statistical issues relevant to the evaluation of efficacy in detail. However, they are not very informative about how to address the evaluation of safety. Guideline E1 (Population exposure for assessing safety of chronic treatments for non-life-threatening conditions) provides recommendations as to the number of patients to treat (300–500) for at least 6 months in prospective studies and the number to treat for at least a year (100 patients or more). Guideline E2 (Safety data management: Expedited reporting and individual case reports) provides specific definitions of “adverse event,” “adverse drug reaction,” “unexpected drug reaction,” and “serious adverse event or drug reaction,” but no guidance on statistical design or analysis issues. Guideline E2E (Pharmacovigilance planning) briefly mentions methods for evaluating spontaneous reports, including Bayesian methods and data mining techniques for evaluating drug–drug interactions, but does not provide details. The guideline also includes a brief summary of epidemiologic methods that could be useful for evaluating adverse events. Guideline E3 (Structure and content of clinical study reports) provides some general comments, but no specific recommendations about evaluating safety. For example, when discussing exposure, “the more common adverse events, laboratory test changes etc. should be identified, classified in some reasonable way, compared for treatment groups, and analysed, as appropriate, for factors that may affect the frequency of adverse reactions/events” (p. 19). There appears to be a recognition that some focus generally is well advised; for example, “[i]t is not intended that every adverse event be subjected to rigorous statistical evaluation” (p. 22). Guideline E5 (Ethnic factors), which is primarily concerned with bridging studies, recommends evaluating rates of common adverse events in an efficacy bridging study or a separate safety study (p. 6), with no recommendations as to appropriate designs or analysis methods. Guideline E9 (Statistical principles for clinical trials) addresses safety only briefly (pp. 28–29), recommending that comprehensive safety and tolerability measures should be collected, including type, severity, onset, and duration of adverse events. However, it is “not always self-evident how to assess incidence” and, in fact, “[i]n most trials the safety and tolerability implications are best addressed by applying descriptive statistical methods to the data, supplemented by calculation of confidence intervals whenever this aids interpretation” (p. 29). The guideline comes closest to recommending specific statistical analysis approaches on p. 32 when discussing the evaluation of information from a safety database: “The evaluation should also make appropriate use of survival analysis methods to exploit the potential relationship of the incidence of adverse effects to duration of exposure and/or follow-up. The risks associated with identified adverse effects should be appropriately quantified to allow a proper assessment of the risk/benefit relationship.” Guideline E14 (The clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs) provides some general recommendations, but no methodologic specifics. There is a need for a negative thorough QT/QTc study showing that the upper bound of a 95% one-sided confidence interval excludes 10 ms, and the guideline recommends carrying out analyses using uncorrected ECG values and values corrected using Bazett, Fridericia, and possibly linear approximation corrections to adjust for heart rate differences.

It is clear that there is a need for more definitive guidance on the use of statistical methods for assessing safety, for designing trials to provide useful clinical perspective about safety, and for making most effective use of new technologies that may be appropriate for very early identification of potential toxicity issues. The objective of this book is to provide recommendations and guidance for the effective application of statistical methods, both old and new, to the evaluation of drug and other medical product safety.

## References

- [1] Olsen, A.K. and Whalen, M.D. (2009) Public perceptions of the pharmaceutical industry and drug safety. *Drug Safety*, **32**, 805–810.
- Surveys conducted in 2006–2007 identified concerns about the effectiveness of drug companies and the FDA in assuring drug safety during development and after approval. This article addresses reforms aimed at addressing the concerns, including promoting a “safety culture throughout the health system” as a goal of every corporate and organizational entry, bringing stakeholders together to work towards achieving such a safety culture via public–private partnerships such as the FDA Sentinel Network and encouraging all healthcare professionals to track and report adverse events, and identifying a key role for pharmacovigilance professionals in promoting the understanding of the importance of cooperative activity in implementing the activities necessary to assure drug safety.
- [2] Arrowsmith, J. (2011) Phase III and submission failures: 2007–2010. *Nature Reviews Drug Discovery*, **10**, 87–87.
- This article summarizes reasons for failure of compounds during Phase 3 and after submission. The safety failure rate was 21% of 83 products that failed. One conclusion from the entire review: many failures occurred with drugs having novel mechanisms of action in areas of high unmet medical need.
- [3] Arrowsmith, J. (2011) Phase II failures: 2008–2010. *Nature Reviews Drug Discovery*, **10**, 328–329.
- Summary of reasons for failure of compounds during Phase 2; 19% (17 of 87) failed for safety.
- [4] Kola, I. and Landis, J. (2004) Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery*, **3**, 711–715.
- Most pharmaceutical products (about 90%) were observed to fail in development between 1991 and 2000, with some variation by therapeutic area. About 25% were observed to fail during the registration review period, when almost all development costs had been incurred. Major causes of attrition in 2000 were lack of efficacy and safety (toxicology and clinical safety), both accounting for about 30% of failures. The high attrition rate is not economically sustainable for the industry. A number of proposals are presented to address the problem, including implementing testing methods that can effectively eliminate compounds with potential toxicity, especially mechanism-based toxicity, early in the development process.
- [5] Walker, D.K. (2004) The use of pharmacokinetic and pharmacodynamic data in the assessment of drug safety in early drug development. *British Journal of Clinical Pharmacology*, **58**, 601–608.
- It is important to assure that the target therapeutic free drug concentration required for efficacy, which often can be estimated from in vitro studies, is likely to be achieved by a realistic dosage regimen. This is extendable to assess potential safety risk when plasma free concentration is predictive of compound safety. However, many limiting safety findings, for example, hepatotoxicity, are unrelated to systemic concentrations. Pharmacogenetic considerations also are important, for example, dependence on CYP2D6 or CYP2C19 metabolism may present an undesirable safety risk. Single-dose pharmacokinetic data can provide guidance for deciding about the further development of a compound when there is information relating predictability of free drug concentration for efficacy or safety.
- [6] Williams, R. (2011) Discontinued drugs in 2010: oncology drugs. *Expert Opinion on Drug Discovery*, **20**, 1479–1496.
- This paper reviews the reasons for discontinuation of 28 oncology drugs dropped from the worldwide pipeline in 2010. Most discontinuations were for strategic or unspecified reasons, probably reflecting lack of resources or of evidence that the drug could be differentiated from or represent an improvement over current therapies. Three were dropped for safety-related reasons.
- [7] DiMasi, J.A., Hansen, R.W. and Grabowski, H.G. (2003) The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, **22**, 151–185.

## 12 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

Comprehensive analysis of the costs of drug development for 68 “randomly selected” products from 10 companies, updating a similar study published 12 years earlier. Average estimated total R&D cost is about \$800 million (actually closer to \$900 million counting post-approval R&D cost) as opposed to \$231 million from the previous study. A number of factors contribute to increased cost, including more difficult therapeutic targets and more comprehensive studies. Most of the increase in cost comes at the clinical development as opposed to preclinical stage.

- [8] Qureshi, Z.P., Seoane-Vazquez, E., Rodriquez-Monguio, R. and Stevenson, K.R. (2011) Market withdrawal of new molecular entities approved in the United States from 1980 to 2009. *Pharmacoepidemiology and Drug Safety*, **20**, 772–777.

This paper examines the bases for withdrawal of new molecular entities from the market between 1980 and 2009 based on data from FDA and other sources. During the period 740 NMEs were approved by FDA, and 118 (16%) were withdrawn. Withdrawal rates varied across categories of NME. Safety was the primary reason for withdrawal for 26 products (22% of withdrawals). About half of the discontinuations were for drugs approved in the 1980s, about 40% for drugs approved in the 1990s, and about 10% for drugs approved from 2000 on. Among safety withdrawals, about half occurred for drugs approved in the 1990s, 40% for drugs approved in the 1980s, and about 10% for drugs approved from 2000 on.

- [9] Laverty, H., Benson, C., Cartwright, E.J. *et al.* (2011) How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? *British Journal of Pharmacology*, **163**, 675–693.

This is a report from a workshop hosted by the MRC Centre for Drug Safety Science to address issues retarding cardiovascular safety, focusing on three questions: (a) What are the key cardiovascular safety liabilities in drug discovery and development, and clinical practice? (b) How good are preclinical and clinical strategies for detecting these liabilities? (c) Is there a mechanistic understanding of the liabilities? With respect to (a), most adverse events (AEs) reported in the FDA’s Adverse Event Reporting System often are not well described, nor are the AE–drug relationships established. AEs leading to discontinuation or withdrawal can be rare, for example, less than 10 per million patients for Torsade des Pointes arrhythmias. Myocardial ischemia or necrosis, heart failure, and coronary artery disorders often are not reported early in drug development, suggesting that they are not being captured by preclinical assays or during development. However, drugs with drug-induced vascular injury have been successfully developed. With regard to (b), it was observed that preclinical safety evaluation generally is conducted in young healthy animals lacking the pathophysiological background underlying these disease conditions, so it is not clear whether the usual preclinical models accurately reflect the patient population ultimately to be exposed to drug. QT-related assays appear to be good, but not perfect, predictors of arrhythmia in clinical setting. The picture is much less clear for other AEs such as myocardial ischemia. Certain aspects of drug-induced cardiovascular (CV) disturbances are not routinely addressed during preclinical development. With regard to (c), although the mechanism for QT prolongation is known, arrhythmias account for only a small proportion of drug withdrawals due to CV safety issues. The mechanism for most of these is unclear. Targeted cancer therapies, particularly protein kinase inhibitors, also can lead to CV toxicity because they inhibit kinases necessary for maintaining homeostasis in cardiac tissue, so that risk must be balanced against benefit. Two key points from the report are the need for approaches that discover the fundamental mechanisms of CV liability to allow a step change in the detection of CV liability early in drug development, and that current understanding about mechanisms resulting in CV dysfunction is inadequate.

- [10] Rolan, P., Danhof, M., Sanski, D. and Peck, C. (2007) Current issues relating to drug safety especially with regard to the use of biomarkers: a meeting report and progress update. *European Journal of Pharmaceutical Sciences*, **30**, 107–112.

This paper is a summary of the discussion at a meeting of an expert group to review the state of the art in detecting drug-related safety problems and the role of biomarkers and modeling techniques in improving the ability to detect toxicity issues early in drug development. Without a causal

mechanism linking a biomarker with a clinical endpoint and toxicity it is difficult to distinguish between a biomarker change reflecting exposure and a change reflecting toxicity. Even for hepatotoxic drugs, it is difficult to predict whether a patient demonstrating a change will go on to experience toxicity. There are few useful skin-based biomarkers for prediction/development of toxicity. There are also few new biomarkers of hepatotoxicity; current ones have poor specificity. Hemotoxicity and immunotoxicity are usually easy to detect in development of oncology drugs, less successful for non-cytotoxic and non-immunosuppressive drugs. Many drugs are toxic to bone marrow through non-humoral mechanisms, but current diagnostic tests have poor diagnostic properties. Preclinical models are not very useful for predicting behavioral toxicity. Simple behavioral toxicities like sedation are easy to detect early, but toxicities affecting complex behavior (e.g., depression) are difficult to detect. Modeling biological systems/responses mechanistically may be useful for predicting subtle toxicities.

- [11] Marrer, E. and Dieterle, F. (2010) Impact of biomarker development on drug safety assessment. *Toxicology and Applied Pharmacology*, **243**, 167–179.

This article addresses considerations about the role and objectives of safety biomarkers at each stage of drug development. Effective biomarkers can enable the therapeutic use even of drugs that may be toxic if the toxicity potential can be recognized early and managed clinically. The article provides an overview of safety biomarkers for specific organ systems (kidney, liver, heart, vascular system) along with evidence supporting their use in preclinical and clinical development. Current standards for monitoring renal safety are late in identifying toxicity, insensitive, and not very specific. However, a number of new protein-based biomarkers have promise for early detection of acute kidney injury. The picture is similar for liver toxicity; a number of biomarkers used together may help in early prediction of hepatotoxicity. Cardiac troponins are recently studied biomarkers for damage to heart muscle that have reasonable specificity and sensitivity. Vascular safety is a major concern because of a lack of diagnostic markers and gaps in understanding of pathogenesis and mechanisms of vascular lesion development. The article discusses issues related to stages of development of biomarkers: identification, preclinical qualification, and clinical qualification and diagnostic use.

- [12] Bass, A.S., Cartwright, M.E., Mahon, C. *et al.* (2009) Exploratory drug safety: a discovery strategy to reduce attrition in development. *Journal of Pharmacological and Toxicological Methods*, **60**, 69–78.

This paper describes a number of areas of study aimed at mitigating risk of failure during development, with primary emphasis on preclinical and early clinical development stages. These include pre-development safety pharmacology to ascertain pharmacodynamic properties of test molecules on major organ system function, genetic toxicology to evaluate potential genotoxicity of active constituents, metabolites, and excipients, exploratory drug metabolism and pharmacokinetic studies, studies aimed at evaluating potential for off-target activity of a test molecule that could present potential safety liabilities. Also important are exploratory drug safety studies aimed at identifying toxic effects that may be evident in extended administration (up to 14 days), to identify issues that would be likely to emerge in longer-term chronic studies.

- [13] Benbow, J.W., Aubrecht, J., Banker, M.J. *et al.* (2010) Predicting safety toleration of pharmaceutical chemical leads: cytotoxicity correlations to exploratory toxicity studies. *Toxicology Letters*, **197**, 175–182.

Cytotoxicity assessments are used to conduct high-throughput safety screening for evaluating compounds considered for further development. Difficulties in extrapolating *in vitro* cytotoxicity to *in vivo* effects have been reported due to direct correlations or limitations caused by confounding factors of the whole organism. Moreover, relationships found on a limited set of compounds may not apply for a wider set of compounds. This article addresses the hypothesis that compounds with less cytotoxic potential would have fewer safety findings in short-term rat exploratory toxicity studies using a wide range of pharmaceutically relevant compounds (72 compounds). A composite safety score was generated for each compound based on the incidence and severity of adverse

## 14 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

outcomes using a scoring algorithm based on systemic toleration, organ functional assessment, and multiorgan pathology. Results indicate that a simple cytotoxicity assessment can be a useful addition to the battery of information usually obtained during lead development.

- [14] Merlot, C. (2010) Computational toxicology – a tool for early safety evaluation. *Drug Discovery Today*, **15**, 16–22.

This insightful review focuses on recent developments in computational toxicology aimed at predicting the toxicity of a molecule from a representation of its chemical structure. Algorithms for the computations mostly invoke expert systems or statistical modeling. There are a number of key issues associated with either approach. One of these is the reliability of a prediction for one or more compounds, and much effort has been devoted to defining the applicability of computational models. The methods work best when the test compounds are similar to the compounds used to train the model; failures of the approaches can be due to differences between the test compounds and the set of training compounds. However, even being close to the training set does not guarantee correct prediction because small structural differences can have substantial toxicological effects. Complex endpoints are difficult to predict, and carcinogenicity, liver toxicity, and developmental toxicity are complex endpoints whose prediction by current models is inadequate. The role of computational toxicology is to shift compound attrition early in the development cycle, to fail cheaply. While computer screens can identify many false positives, they can be used as a preliminary screening device and followed up by *in vitro* or *in vivo* testing.

- [15] Nigsch, F., Macaluso, N.J.M., Mitchell, J.B.O. and Zmuidinavicius, D. (2009) Computational toxicology: an overview of the sources of data and of modelling methods. *Expert Opinion on Drug Metabolism & Toxicology*, **5**, 1–14.

Computational toxicology presents an opportunity to save time and expense in evaluating the toxic potential of drugs and their metabolites, but is limited by the need for comprehensive and extensive data in structured, machine-accessible form. Included within this purview is the emerging field of toxicogenomics driven by the possibility of extensive genome-wide expression analyses, although its practical application remains in doubt. The goals are much clearer than the current ability to achieve them. The article provides an overview of commonly used toxicity assays and discusses a number of kinds of computational approaches, including quantitative structure–activity relationships, target prediction models, and protein target and structure-based methods. Computational methods usually are knowledge-based, relying on a substantial body of expert opinion, or statistically based, relying on substantial data mining activity.

- [16] Nigsch, F., Lounkine, E., McCarren, P. *et al.* (2011) Computational methods for early predictive safety assessment from biological and chemical data. *Expert Opinion on Drug Metabolism & Toxicology*, **7**, 1497–1511.

Predictive safety assessment is incorporated at various stages of drug development in a number of ways: *in silico* methods are used in early discovery to prioritize compounds or to flag potential liabilities; *in vitro* assays are performed to filter out compounds with toxic potential; following lead selection, animal experiments are performed in various species to evaluate pharmacokinetics, pharmacodynamics, potential biomarkers toxicology, gene expression, etc. Authors suggest that the role of computational toxicology is to identify potential issues needing to be followed up by more conventional safety assessment techniques. This article describes different categories of computational toxicology tools: expert- or rule-based systems, statistical models, quantum mechanical calculations, structure-based approaches, and the use of safety panels. Computational approaches may be able to help uncover associations of AEs with chemical structure and activity. Bayesian models have been used to identify relationships between *in vivo* binding profiles and AEs.

- [17] Ioannidis, J.P.A. and Lau, J. (2002) Improving safety reporting from randomized trials. *Drug Safety*, **25**, 77–84.

Empirical evidence across diverse medical fields suggests that the reporting of safety information in clinical trials is largely neglected and receives less attention compared with efficacy

outcomes. Safety data need to be collected and analyzed in a systematic fashion and active surveillance for toxicity during the conduct of a randomized trial is preferable to passive surveillance. Common errors include (1) not reporting safety data at all; (2) making only vague statements; (3) reporting events without a breakdown by study arm; (4) lumping different kinds of AEs under broad categories; (5) combining severity levels; (6) giving  $p$ -values but no event counts; (7) providing information on only a few or the most common AEs; (8) not providing information on AEs leading to withdrawal from the trial; (9) over-interpreting and over-analyzing safety data; (10) over-interpreting the absence of AEs; (11) failing to define scales used to categorize AE severity; (12) reporting data without relevant information about the experimental unit, such as duration of exposure. Some recommendations from the article are as follows: (1) specify the number of patients withdrawn because of AEs per study arm and AE; (2) use widely known, standardized scales for AEs; (3) specify the schedule for safety information collection, specific tests performed, questionnaires used, and whether surveillance was active or passive; (4) provide the number of specific AEs per study arm and per type of AE; (5) tabulate safety information per study arm and severity grade for each AE.

- [18] Ioannidis, J.P.A., Evans, S.J.W., Gotzsche, P.C. *et al.* (2004) Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of Internal Medicine*, **141**, 781–788.

This article describes an extension of 10 new recommendations to the standard CONSORT checklist about reporting harms-related issues. The recommendations are as follows: (1) State in title or abstract if the study collected data on harms and benefits. (2) The introduction should state if the trial addresses both harms and benefits. (3) Define the recorded AEs in the Methods section, clarifying whether all AEs or only a selected sample are included, whether only expected or also unexpected AEs are included, etc. (4) Clarify how harms-related information was collected. (5) Describe statistical methods planned for presenting and analyzing harms information. (6) Describe the participant withdrawals due to harms from each study arm, and the experience with the allocated treatment. (7) Provide the denominators for analyses of harms. (8) Present the absolute risk of each AE (type, grade, severity) per arm, and present appropriate metrics for recurrent events. (9) Describe any subgroup analyses and exploratory harms analyses. (10) Provide in the Discussion section a balanced discussion of benefits and harms, with emphasis on study limitations, generalizability, etc.

- [19] Ioannidis, J.P.A., Mulrow, C.D. and Goodman, S.N. (2006) Adverse events: the more you search, the more you find. *Annals of Internal Medicine*, **144**, 298–300.

Medication-related harms can be identified from many sources, typically case reports, observational studies, and randomized trials. How one defines and looks for problems affects the numbers of AEs that patients report. Patients' judgments about tolerable harm can depend on whether they felt they had effective therapeutic alternatives. It is important to follow appropriate guidelines specifying how and when harms-related information was collected. It is almost always inappropriate to make statements about no difference in AE rates based on non-significant  $p$ -values.

- [20] Balakrishnan, N., Read, C.B., Vidakovic, B. *et al.* (2010) *Methods and Applications of Statistics in the Life and Health Sciences*, John Wiley & Sons, Inc., Hoboken, NJ.
- [21] Chow, S.C. and Liu, J.P. (2013) *Design and Analysis of Clinical Trials: Concepts and Methodologies*, 3rd edn, John Wiley & Sons, Inc., Hoboken, NJ.
- [22] Everitt, B.S. and Palmer, C.R. (2009) *The Encyclopaedic Companion to Medical Statistics*, John Wiley & Sons, Inc., Hoboken, NJ.
- [23] Gad, S.C. (2009) *Drug Safety Evaluation*, John Wiley & Sons, Inc., Hoboken, NJ.
- [24] Lachin, J.M. (2000) *Biostatistical Methods: The Assessment of Relative Risks*, John Wiley & Sons, Inc., New York.
- [25] Moyé, L.A. (2006) *Statistical Monitoring of Clinical Trials: Fundamentals for Investigators*, Springer, New York.

- [26] Proschan, M.A., Lan, K.K.G. and Wittes, J. (2006) *Statistical Monitoring of Clinical Trials*, Springer, New York.
- [27] Rao, C.R., Miller, J. and Rao, D.C. (2007) *Epidemiology and Medical Statistics*, Handbook of Statistics, vol. 27, Elsevier, New York.
- [28] Senn, S. (2008) *Statistical Issues in Drug Development*, 2nd ed., John Wiley & Sons, Ltd, Hoboken, New Jersey.
- [29] Popper, K.R. (1963) Science as falsification, in *Conjectures and Refutations*, Routledge & Kegan Paul, London.

This is a short, elegant, and clear explanation of what it takes for a theory to be “scientific.” Some points: (1) Confirmations or verifications can be found for any theory if they are sought. (2) Confirmations should count only if they result from “risky predictions”, i.e., if there are possible outcomes that could not be explained by the theory. (3) The more a theory forbids, the better it is. (4) A theory which is not refutable is not scientific. (5) Every genuine test of a theory is an attempt to falsify or refute it. (6) Confirming evidence should count only when it can be presented as a serious but unsuccessful attempt to falsify the theory. In short, the scientific status of a theory is determined by its falsifiability, refutability, or testability. The criteria for refutation have to be laid down before obtaining observations to test a theory, i.e., the hypothesis must be stated before the data are obtained.

- [30] Altman, D.G. and Bland, J.M. (1991) Improving doctors’ understanding of statistics. *Journal of the Royal Statistical Society, Series A*, **154**, 223–267.

This article discusses the need for statistical knowledge in medicine, what doctors need to know about statistics and whether the state of statistical knowledge among doctors is adequate. The article considers a number of ways in which doctors acquire statistical knowledge, such as undergraduate and postgraduate education, the quality of many textbooks, and the examples given by papers in medical journals. A number of recommendations are offered, including the following: (1) greater emphasis on statistical principles in undergraduate medical education and more postgraduate statistics courses for doctors, taught by experienced medical statisticians; (2) greater attention to the scientific and statistical correctness of papers published by medical journals; (3) more involvement of statisticians in medicine, at all levels of teaching, refereeing medical papers, membership on ethical committees, and more collaboration with doctors and statistical consultancy. The article is accompanied by extensive commentary by many statisticians.

- [31] Windish, D.M., Huot, S.J. and Green, M.L. (2007) Medicine residents’ understanding of the biostatistics and results in the medical literature. *Journal of the American Medical Association*, **298**, 1010–1022.

This article describes the result of a survey completed by 277 internal medicine residents in 11 residency programs aimed at evaluating their understanding of biostatistics and research result interpretation. The instrument used for the evaluation was a biostatistics/study design multiple-choice knowledge test. The overall mean percentage correct on statistical knowledge and interpretation of results for the residents was 41% vs 72% for fellows and general medicine faculty with research training. Higher scores were associated with additional advanced degrees (50% vs 41%), prior biostatistics training (45% vs 38%), enrollment in a university-based training program (43% vs 36%), or being male (44% vs 39%). Although most (82%) correctly interpreted a relative risk, the residents were less likely to know how to interpret an adjusted odds ratio from a multivariate regression analysis (37%) or the results of a Kaplan–Meier analysis (10%). Most (75%) did not believe that they understood all of the statistics they encountered in journal articles, but almost all (95%) felt it was important to understand these concepts to be an intelligent reader of the literature. The authors conclude that most residents in this study lacked the knowledge in biostatistics needed to interpret many of the results in published clinical research and that residency programs should include more effective biostatistics training in their curricula.



- [32] Wulff, H.R., Andersen, B., Brandenhoff, P. and Guttler, F. (1987) What do doctors know about statistics? *Statistics in Medicine*, **6**, 3–10.

A 10-item multiple-choice test to evaluate statistical knowledge was sent to 250 Danish doctors selected at random from a registry of Danish physicians, who were asked to complete the test without consulting a textbook of statistics; 140 of the doctors completed the questionnaire, as did an additional 97 participants in postgraduate courses in research methods. The median number of correct answers was 2.4 among the random sample and 4 among the additional cohort. The conclusion was that the statistical knowledge of most doctors was so limited that they could not be expected to interpret statistical findings in medical journals correctly.

- [33] Food and Drug Administration (2010) The Sentinel Initiative, <http://www.fda.gov/Safety/FDASSentinelInitiative/default.htm> (accessed 22 July 2014).

An update on FDA progress in building a national electronic system for monitoring post-marketing safety of FDA-approved drugs and other medical products.

- [34] ICH Expert Working Group (1996) Guideline for Good Clinical Practice E6(R1), [http://www.ich.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Efficacy/E6\\_R1/Step4/E6\\_R1\\_Guideline.pdf](http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6_R1/Step4/E6_R1_Guideline.pdf) (accessed 22 July 2014).

A comprehensive international ethical and scientific quality standard for designing, conducting, recording, and reporting trials that involve the participation of human subjects. The objective of the guideline is to provide a unified standard for the EU, Japan and the USA to facilitate the mutual acceptance of clinical data by the various regulatory agencies.

- [35] *ICH Expert Working Group* (2014) ICH Safety Guidelines, <http://www.ich.org/products/guidelines/safety/article/safety-guidelines.html> (accessed 22 July 2014).

- [36] Aboelsoud, N.H. (2010) Herbal medicine in ancient Egypt. *Journal of Medicinal Plants Research*, **4**, 82–86.

This paper summarizes documentary evidence that therapeutic herbs and foods were used extensively in ancient Egypt. Medicines also were made from mineral substances. Medications were age-specific. Prescriptions were written with high skill. Most of the medical knowledge was set by 2000 BCE.

- [37] Anon. Bald's Leechbook, [http://en.wikipedia.org/wiki/Bald's\\_Leechbook](http://en.wikipedia.org/wiki/Bald's_Leechbook) (accessed 13 September 2011).

This Old English medical text in two volumes was probably compiled in the ninth century CE. The first book deals with external disorders, the second with internal problems.

- [38] Anon. Authentic Mayan Medicine, [http://www.authenticmayan.com/maya\\_medicine.html](http://www.authenticmayan.com/maya_medicine.html) (accessed 12 September 2011).

This provides an overview of principles of Mayan medicine. Mayan medicine was holistic by nature, incorporating a medico-religious tradition that took account of the emotional as well as the physical state of the patient. There was an extensive, largely herbal-based, pharmacopoeia.

- [39] Jia, W., Gao, W.-Y., Yan, Y.-Q. *et al.* (2011) The rediscovery of ancient Chinese herbal formulas. *Phytotherapeutic Research*, **18**, 681–686.

This review of ancient Chinese herbal medicine principles describes the organization of formulas containing combinations of herbs with specific roles (monarch, minister, adjuvant, guide). It gives examples where the properties/effects of a particular herbal component can be modified, sometimes dramatically, by the inclusion of other herbal components. The mechanism of action of most combination formulas is still unknown. It discourages use of herbal medicines chronically or in high dosages without the involvement of a skilled practitioner diagnosis and the determination of a holistic treatment approach.

- [40] Lev, E. (2002) Traditional healing with animals (zootherapy): medieval to present-day Levantine practice. *Journal of Ethnopharmacology*, **85**, 107–118.

This paper reviews the history of the use of medicines derived from animal bodies and organs in the Levant. It provides a detailed summary of names and references to documents from various

## 18 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

periods of 99 substances of animal origin identified as being used in traditional medicine from early medieval to present times. Main animal sources include honey, wax, adder, beaver testicles, musk oil, coral, and ambergris. It also provides a detailed list of some 77 animal products currently used.

- [41] Nissenbaum, A. (1993) The Dead Sea – an economic resource for 10,000 years. *Hydrobiologia*, **267**, 127–141.

This paper describes various products of the Dead Sea since ancient times. Medicinal use was made of distilled asphalt for treating skin diseases. Dead Sea waters may have been useful for treating eye disease, and are still used for treating psoriasis.

- [42] Rahman, S.Z., Khan, R.A. and Latif, A. (2008) Importance of pharmacovigilance in Unani system of medicine. *Indian Journal of Pharmacology*, **40**, 17–20.

This paper provides a historical account of pharmacology over 1000 years ago, when more than 2000 drugs were known and studied. It includes a description of the encyclopedic work of Ibn Sinâ (980–1037 CE) who wrote treatises on cardiac drugs, pharmacologic and pharmacotherapeutic characteristics and methods of preparation of many compounds, and many aspects of toxicology. The notion of patient-tailored therapy also appears in the writings of Ibn Sinâ. The paper also discusses early pharmacovigilance and its relation to Unani (Greco-Arabic) medicine. It recognizes the present-day need for systematic data on the incidence of adverse events associated with the use of traditional medicines.

- [43] Saad, B., Azaizeh, H. and Said, O. (2005) Tradition and perspectives of Arab herbal medicine: a review. *Evidence Based Complementary and Alternative Medicine*, **2**, 475–479.

This is an overview of the history of Arab medicine. Major innovations include the discovery of the immune system and introduction of microbiological science; the separation of medicine from pharmacological science (Avicenna, tenth and eleventh centuries CE); and advances in herbal medicine including extraction of anesthetic compounds from local herbs and the introduction of 350 new plant species to medicinal herbs.

- [44] Voultziadou, E. (2010) Therapeutic properties and uses of marine invertebrates in the ancient Greek world and early Byzantium. *Journal of Ethnopharmacology*, **130**, 237–247.

Review of ancient Greek, Roman, and Byzantine texts regarding the therapeutic properties of marine invertebrates. For over 30 species, provides the scientific, classical, and common names, summarizes their therapeutic properties and uses, and provides references to original texts.

- [45] Merchant, B. (1998) Gold, the noble metal and the paradoxes of its toxicology. *Biologicals*, **26**, 49–59.

This paper describes therapeutic uses of metallic gold and gold salts, and some related toxicity issues.

- [46] Sarkar, P.K. and Chaudhary, A.K. (2010) Ayurvedic bhasma: the most ancient application of nanomedicine. *Journal of Scientific and Industrial Research*, **69**, 901–905.

Sarkar and Chaudhary describe the preparation of nanoparticle formulations of various metals for therapeutic use.

- [47] Griffin, J.P. (2004) Venetian treacle and the foundation of medicines regulation. *British Journal of Clinical Pharmacology*, **58**, 317–325.

This paper describes the history of legislation from the fifteenth century onward regarding oversight of the preparation and assurance of quality of a “universal panacea” known as Venetian treacle. This early legislation stimulated concerns about the quality of all medicines and was the earliest implementation of medicine regulation.

- [48] Chan, T.Y.K. (1994) The prevalence, use, and harmful potential of some Chinese herbal medicines in babies and children. *Veterinary and Human Toxicology*, **36**, 238–240.

This article reviews the prevalence use of Chinese herbal medicines (CHM) in Chinese pregnant women, babies and children living in Hong Kong and the harmful potential of some CHM and Chinese proprietary medicines (CPM) in babies and children. The use of CHM appears to be common amongst Chinese pregnant women. The possible effects of these herbs on the fetus

and baby and their overall safety are not known. This practice should be discouraged since there is a suggestion that maternal consumption of CHM might increase the risk of neonatal jaundice. Both *chuen-lin* and *yin-chen* can displace bilirubin from their serum protein binding and increase the risk of hyperbilirubinemia. These herbs should not be given to the neonates. The use of CPM containing undeclared drugs of high toxicity or lead, arsenic and mercurial compounds should be banned. The medical profession and the general public should be alerted to the harmful potential of some CHM and CPM. There should be continuing efforts to collect information on the safety of these compounds.

- [49] Chitturi, S. and Farrell, G.C. (2008) Herbal hepatotoxicity: an expanding but poorly defined problem. *Journal of Gastroenterology and Hepatology*, **15**, 1093–1099.

Many herbal remedies are hepatotoxic, and many more may be hepatotoxic to an unknown degree. This article provides a list of hepatotoxic herbal compounds and provides several examples of how the hepatotoxicity is manifested.

- [50] Ernst, E. (2003) Serious psychiatric and neurological adverse effects of herbal medicines – a systematic review. *Acta Psychiatrica Scandinavica*, **108**, 83–91.

This is a survey of literature, mostly case reports, of adverse events associated with the use of alternative (mostly herbal) medicines. The reports suggest that herbal medicines have often been associated with potentially severe psychiatric and neurological adverse events. The article summarizes reports of herb–drug interactions and adverse events associated with contamination or adulteration of the herbal products. The findings are suggestive because many of the reports lack the level of detail necessary to establish clear causal or even associative relationships. However, the reports do suggest that the possibility of herb–drug pharmacokinetic or pharmacodynamic interactions needs to be taken into account in patient management. Since herbal medicines (at least those mentioned in the article) have not been rigorously evaluated for efficacy, informed risk–benefit assessments rarely can be made.

- [51] Hu, Z., Yang, X., Ho, P.C.L. *et al.* (2005) Herb–drug interactions. A literature review. *Drugs*, **65**, 1239–1282.

Herb–drug interactions probably are significantly under-reported and underestimated for various reasons. For example, patients do not tell their physicians about the herbal remedies they are using; there is a lack of regulations requiring rigorous preclinical/clinical assessment of herbal remedies; most clinical trials of herbal remedies are poorly designed and executed; there is no comprehensive AE surveillance system in many countries (especially those where herbal remedies are popular); and any herbal compound may contain multiple bioactive constituents whose individual actions are difficult to separate. This article is a substantial survey of literature (with more than 500 references) on interactions between herbal preparations and a wide range of conventional pharmaceutical products.

- [52] Niggemann, B. and Grüber, C. (2003) Side-effects of complementary and alternative medicine. *Allergy*, **58**, 707–716.

This article is an extensive literature review of mostly anecdotal reports of various kinds of adverse events associated with herbal medicines and mechanical procedures such as acupuncture. It states that “[v]irtually all herbal remedies have been reported to cause either allergic sensitization or photosensitization.” Organ toxicity, especially hepatotoxicity and nephrotoxicity, of herbal medicines can be due to the chemical constituents of the product, or to contamination, adulteration, or misidentification of ingredients. Chronic use of high doses of some herbal preparations has been associated with increased cancer risk.

- [53] Saad, B., Azaizeh, H., Abu Hijleh, G. and Said, O. (2006) Safety of traditional Arab herbal medicine. *Evidence Based Complementary and Alternative Medicine*, **3**, 433–439.

This is a systematic review of the safety of traditional Arab medicine and contributions of Arab scholars to toxicology. Ancient Arab sources on toxicology go back to the eighth to tenth centuries

## 20 SAFETY IN MEDICAL PRODUCT DEVELOPMENT

CE. The article summarizes toxicologic findings for commonly used traditional herbal medicines. It provides recommendations for processes to assure safety and consistency of herbal medicines.

- [54] Skoulidis, F., Alexander, G.J.M. and Davies, S.M. (2005) Ma huang associated acute liver failure requiring liver transplantation. *European Journal of Gastroenterology and Hepatology*, **17**, 581–584.

This is a case report of hepatotoxicity requiring liver transplant induced by ma huang, which contains ephedrine-type alkaloids. It also reviews literature about potential ma huang induced hepatotoxicity.

- [55] Wojcikowski, K., Johnson, D.W. and Gobé, G. (2004) Medicinal herbal extracts – renal friend or foe? Part one: the toxicities of medicinal herbs. *Nephrology*, **9**, 313–318.

Literature review of articles describing cases of renal toxicity.

- [56] Lee, P.R. and Herzstein, J. (1986) International drug regulation. *Annual Review of Public Health*, **7**, 217–235.

Drug regulation has a long history, dating back 3000 years. Apothecaries were regulated in Europe and Muslim countries in the Middle Ages, with focus shifting to regulations of drugs from the sixteenth century CE on. The institution of patent laws in the USA and Europe and the isolation of pure morphine in 1805 set the stage for the evolution of the pharmaceutical industry and subsequent development of international drug regulation.

- [57] Levey, M. (1963) Fourteenth century Muslim medicine and the *hisba*. *Medical History*, **7**, 176–182.

The article describes the office of the *hisba*, established in the ninth century CE to enforce regulations regarding public safety (among other things). It quotes extensively from a fourteenth-century text regarding regulations concerning physicians, ophthalmologists, surgeons, and bonesetters, and extensive discussion of preparation of various therapeutic agents, typically herbal

- [58] Swann, J. P. (2014) About FDA: FDA's Origin, <http://www.fda.gov/AboutFDA/WhatWeDo/History/Origin/ucm124403.htm> (accessed 22 July 2014).

This describes the history of the FDA from 1862 through 2001, including the various legislative acts that defined and subsequently expanded its role and responsibility.

- [59] Lee, P.R., Lurie, P., Silverman, M.M. and Lydecker, M. (1991) Drug promotion and labeling in developing countries: an update. *Journal of Clinical Epidemiology*, **44** (Suppl. II), 49S–55S.

This paper provides a description of changes in promotional practices by multinational and domestic pharmaceutical companies as of late 1980s. Legislation to control drugs was enacted in England in the late nineteenth century, in Switzerland in 1900, in the USA in 1906. Regulations relevant to safety and efficacy were enacted in Norway and Sweden in the 1920s, in the USA in 1938, and in virtually all European countries by early 1960s. The paper points to the double standard of basing promotion in developed countries on scientifically based evidence. It says that the practice of expanding indications in developing nations persists, and unjustified claims of efficacy or safety continue(d) to proliferate.

- [60] Food and Drug Administration. About FDA: Significant Dates in US Food and Drug Law History, <http://www.fda.gov/AboutFDA/WhatWeDo/History/Milestones/ucm128305.htm> (accessed 1 September 2011).

Annotated calendar of dates of significant events in FDA history, especially key court cases and legislation.

- [61] Wax, P.M. (1995) Elixirs, diluents, and the passage of the 1938 federal food, drug and cosmetic act. *Annals of Internal Medicine*, **122**, 456–461.

This paper gives a history of the events associated with the Elixir Sulfanilamide disaster of 1937 in which 105 patients died from the use of a preparation of sulfanilamide that used diethyl glycol as a diluent. The events and the subsequent public attention led to the 1938 Federal Food, Drug, and Cosmetic Act that required for the first time testing of new drug products for toxicity.



INTRODUCTION 21

- [62] Council for International Organizations of Medical Sciences, [www.cioms.ch](http://www.cioms.ch) (accessed 8 September 2011).  
Website for CIOMS, World Health Organization.
- [63] International Conference on Harmonisation, <http://www.ich.org/> (accessed 8 September 2011).  
ICH website.
- [64] Safety Guidelines, [http://www.ICH.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Safety/Safety\\_Guidelines.zip](http://www.ICH.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/Safety_Guidelines.zip) (accessed 1 June 2010).  
Archive file containing the latest version of the ICH safety guidelines.
- [65] Efficacy Guidelines, [http://www.ICH.org/fileadmin/Public\\_Web\\_Site/ICH\\_Products/Guidelines/Efficacy/Efficacy\\_Guidelines.zip](http://www.ICH.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/Efficacy_Guidelines.zip) (accessed 1 June 2010).  
Archive file containing the latest version of the ICH efficacy guidelines.

