CC/2016/15

COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

First draft statement on COC/G07: Part d) Alternative testing strategies for carcinogens incorporating results from short-term tests

Attached is the first draft of this Guideline Statement. This is based on the overview paper, CC/2016/07, which was discussed at the July 2016 COC meeting. Paragraphs relating to high-throughput screening technologies (e.g. the 'Tox21' project) have been omitted as this aspect is now included as part of discussion paper CC/2016/14 for G07 part c. Preliminary conclusions have been added in Section 6, based on the Committee's comments and views expressed during discussion of this topic at the meeting in July.

Questions for the Committee

i. Members are invited to comment on the content of the draft Guideline Statement.

ii. Do Members have any comments on the draft Conclusions?

Secretariat
November 2016

Alternatives to the 2-year Bioassay

First draft statement on COC/G07: Part d) ***Alternative testing strategies for carcinogens incorporating results from short-term tests***

Abbreviations

FDA	Food and Drug Administration
IARC	International Agency for Research on Cancer
IATA	integrated approach to testing and assessment
ICH	International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
JPMA	Japanese Pharmaceutical Manufacturers Association
NEG CARC	Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathological Risk Factors for Carcinogenicity
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
PhRMA	Pharmaceutical Research and Manufacturers of America

1. Introduction

D1. The current paradigm for assessing the carcinogenicity of a chemical is a combination of genotoxicity testing and 2-year rodent bioassays. Genotoxicity/mutagenicity tests provide a useful screen to indicate positive carcinogenic potential for those compounds with a genotoxic mode of action, however they produce high numbers of false positive results, have little or no capacity to identify non-genotoxic carcinogens, and generally lack dose–response characterisation. Over the last few decades, the 2-year rodent bioassay has become the gold standard to assess carcinogenicity. Nevertheless there are limitations in extrapolating from the 2-year bioassay to human cancer risk; the primary disadvantage being that positive findings for carcinogenicity in rodents may be of limited or no relevance to human cancer risk due to issues of dose, species specificity and/or mode of action.

D2. New strategies are being developed to assess carcinogenicity in which the use of 2-year rodent bioassays is reduced or replaced with shorter term study data (from *in vivo*, *in vitro* and/or *in silico* tests). Some of these strategies have followed an approach of attempting to predict the outcome of 2-year rodent bioassays, with subsequent evaluation of the applicability of these findings to the human situation, while other approaches aim more directly at identifying and/or assessing the potential for carcinogenicity in humans. As these new strategies are currently in development, they have not been fully validated.

D3. The following section of this guidance statement is a review of developments to date and proposed strategies for future developments with relevance to this topic. The aim of the COC guidance will be to list some of the alternative approaches that are being developed, to discuss some of the advantages and disadvantages of using these approaches, and to consider the potential utility of such approaches for evaluating carcinogenic risks posed to the public from exposure to chemicals present at ambient levels in the environment.

2. Purpose of the Assessment

D4. Carcinogenicity studies are performed for a variety of reasons. These include hazard identification, hazard characterisation, and/or safety assessment of substances such as pharmaceutical products, industrial chemicals, food additives, cosmetics, and chemicals present in the general environment.

D5. The first step in a carcinogenicity assessment is normally a genotoxicity test battery. A lifetime rodent bioassay may then be required depending on the regulatory and legislative setting. For example, for small molecule pharmaceuticals intended for

continuous use or regular intermittent use, data from at least one 2-year rodent carcinogenicity bioassay are currently required by regulatory agencies. Conversely, the use of data from tests performed in vivo is not permitted for some products, such as the constituents of cosmetics intended for sale in the European Union. Regulatory frameworks for carcinogenicity testing of chemicals vary, but in many cases in vivo carcinogenicity bioassays are not performed and/or requested.

D6. For public health protection relating to chemicals present at ambient levels in the environment, the principal goals of carcinogenicity evaluations are the identification and risk assessment of human-relevant carcinogens. The aim is to decide whether exposure to a certain level of a particular chemical is acceptable in terms of the likelihood that it will cause cancer in humans, and to allow for management of this risk. The task is complex as the answer required is not a simple, binary 'yes' or 'no' but ideally a probabilistic evaluation of the risk effectively encountered by humans. It also depends on the cancer mode of action for the chemical.

D7. For application to the risk assessment of chemicals present in the environment, new systems for carcinogenicity evaluation would ideally have the potential to produce organ-specific, dose-dependent information relevant to humans.

3. History and Developments To Date

D8. The utility of short-term toxicological findings *in vivo* as an element to predict the outcomes of 2-year rodent carcinogenicity bioassays has been tested in several retrospective analyses of information in existing toxicological databases. Some studies have looked at the ability of short-term findings to predict tumourigenicity at the organ-specific level, whilst others have used a broader approach to evaluate whether the presence or absence of changes in short-term tests can predict tumourigenicity more generally at the whole-organism level.

3.1 Evaluations of the National Toxicology Program (NTP) database

D9. Allen et al. (2004) reviewed existing data in the NTP database with the aim to correlate specific hepatocellular pathology in pre-chronic studies (≤ 12 months) with carcinogenic endpoints in 2-year studies. Data were evaluated for mice (83 compounds) and rats (87 compounds). The pre-chronic endpoints evaluated were hepatocellular cytomegaly, hepatocellular necrosis, bile duct hyperplasia, hepatocellular hypertrophy, and hepatocellular degeneration (rats only). Increased liver weight was also included. Hepatocellular hypertrophy was the single most predictive feature (10/27 mouse carcinogens, 0 false positives; 5/11 rat carcinogens, 10 false positives). Three features as a group (hepatocellular necrosis, hypertrophy, and cytomegaly) correctly predicted carcinogenicity findings at 2 years for 17/27 (2

false positives) mouse and 7/11 (16 false positives) rat liver carcinogens. Inclusion of liver weight as a fourth criterion improved the sensitivity of the screen, but decreased the specificity (25/27 mouse carcinogens, 18 false positives; 11/11 rat carcinogens, 32 false positives). Genotoxicity results (*Salmonella* test and Micronucleus assay) did not correlate well with liver carcinogenesis outcomes in either mice or rats.

D10. The International Life Sciences Institute (ILSI) Health and Environmental Sciences Institute (HESI) conducted a retrospective analysis of the NTP database to test the hypothesis that the signals of importance for human cancer hazard identification can be detected in shorter term studies than the 2-year bioassay (Boobis et al. 2009). Sixteen chemicals were selected on the basis that they were positive in liver, kidney or lung in lifetime rodent (rat and/or mouse) carcinogenicity bioassays and that genotoxicity and short-term rodent study data were available. Thirteen-week study data for immuno-, liver, kidney and lung toxicity were reviewed for correlation against tumour outcomes in the corresponding tissues in 2-year bioassays.

D11. In *genotoxicity* assays, 5 chemicals were positive, 8 negative and 3 equivocal. The authors noted the requirement for a reliable battery of genotoxicity tests.

D12. Markers of *immune system* changes (downregulation, proliferation, or neoplasia) included haematology (total leukocyte, segmented neutrophil, lymphocyte, and monocyte counts), spleen and/or thymus weights, and histopathological findings in bone marrow, spleen, thymus and lymph nodes. None of the 16 chemicals caused direct immunosuppression in 13-week studies and there was no clear evidence of neoplasia in elements of the immune system. Several chemicals showed immune changes that were attributed to stress. The authors noted the requirement for further definition and evaluation of short-term tests for immunosuppressive effects, suggesting further work to include evaluations using a range of known positive and negative compounds.

D13. Liver findings examined for 13-week studies were organ weight, clinical pathology, and histopathology, including relative liver weight, hepatocellular hypertrophy, altered foci, hepatocyte necrosis, hepatocyte vacuolation, hepatocyte degeneration, bile duct hyperplasia, increased alanine transaminase levels, increased sorbitol dehydrogenase levels, and increased bile acid/bilirubin levels. Six chemicals were tumourigenic in the liver of rats, 9 in mice. In 13-week studies, liver weight was the best single predictor of tumour outcome (5/6 in rat, 6/9 in mouse). Grouping liver weight with other criteria increased the positive predictivity to 6/6 in rat and 8/9 in mouse. Considering the results collectively for rats and mice, there were no false positives, and one false negative (one chemical induced tumours in the mouse bioassay but no changes in rats or mice at 13 weeks). On this basis, the authors concluded that conventional liver endpoints currently identified in 13-week toxicity studies were not adequate to identify all chemicals with carcinogenic

potential and that additional endpoints may identify other key events that might more accurately predict carcinogenic potential in rats and mice. These would then be useful for defining modes of action to assess human carcinogenic potential and risk more effectively. Such endpoints include increases in cell proliferation (S-phase response) and induction/inhibition of apoptosis (measurement of labelling indices for both events), constitutive androstane receptor nuclear receptor activation (reporter assays), cytochrome P450 induction (direct biochemical measurement), and peroxisome proliferation (measurement of palmitoyl coenzyme A oxidase activity). The potential for –omics platforms to identify additional indicators was noted.

D14. Thirteen-week study criteria for *kidney* changes included hyaline droplets, inflammation, chronic progressive nephropathy, and absolute and relative kidney weights. Five compounds were tumourigenic in the kidney of rats, none in mice. All 5 chemicals were positive for 13-week changes. The best predictor of tumourigenicity at 13 weeks was increased kidney weight. When this parameter was combined with histologic findings, no false negatives were identified. The authors noted that 13-week study findings for kidney may give clues to carcinogenicity mode of action, which may help interpretation of human relevance (e.g. 4 chemicals that induced kidney tumours showed increased hyaline droplets, which indicates a rodent-specific mode of action that is not relevant to humans).

D15. For *lung*, diagnostic terms for histomorphologic alterations used by NTP to describe lung lesions in 13-week studies were: chronic active inflammation, inflammation NOS (not otherwise specified), alveolar epithelial hyperplasia, bronchiolar hyperplasia, proteinosis, fibrosis, histiocytic infiltration, and foreign body. In total, 11/16 correct predictions of lung carcinogenesis were made from short-term data. Seven chemicals induced tumours in rats and/or mice. Four of these showed inflammation and/or hyperplasia at 13 weeks and an additional 1 was genotoxic (giving 5 true positives, with the other 2 chemicals being false negatives). Two chemicals induced inflammation and/or hyperplasia at 13 weeks but did not show tumours in the 2-year study, i.e. were false positives, and there were 7 true negatives.

D16. Overall, the authors concluded that for most, but not all, of the chemicals producing tumours in 2-year studies, cellular changes indicative of a tumourigenic endpoint could be identified after 13 weeks using routine evaluations, but that such evaluations are not adequate to identify all non-genotoxic chemicals that will eventually produce tumours in rats and mice. Additional endpoints are needed to identify signals not detected with routine evaluation. Such endpoints might include BrdU labelling and a measure of apoptosis. Further efforts would be required to determine false-positive rates of this approach.

3.2 Pharmaceuticals

D17. Approaches are being developed to allow situations in which the regulatory evaluation of potential human cancer risks from pharmaceuticals may in some cases be made without the requirement for a 2-year rodent bioassay, based on the integration of other data using weight of evidence approaches.

Center for Drug Evaluation and Research (CDER)/FDA

D18. In 1998, the US FDA reviewed the use of 2-year rodent studies and alternative strategies for carcinogenesis testing and stated an aim to move away from reliance on the results of one test (the traditional lifetime bioassay in both sexes of two rodent species) towards a decision-making process based on a profile of data, using a weight of evidence approach that takes into account the increased knowledge of carcinogenic mechanisms that has been gained since the 2-year bioassay was adopted as a routine screen in the 1970s (Schwetz and Gaylor, 1998). A conceptual strategy was proposed, including a preliminary evaluation for genotoxicity to include data on physical–chemical properties, structure alert information, information from computer-based prediction systems and the results of a genetic toxicity screen, and subsequent tests to include transgenic mouse models and then possibly a 2-year rodent study. The inclusion of data relating to non-genotoxic mechanisms of carcinogenicity would be important, including the following mechanisms: hormone modulation, growth factor perturbation, cell proliferation (mitogenic, cytotoxic), inhibition of apoptosis, cell-to-cell communication, P450 induction, spindle fibre effects, altered methylation status, and specific mechanisms (ß-agonist, uterine tissue; H_2 antagonist, glandular stomach; peroxisome proliferation). It was proposed to evaluate these new test systems in parallel with the conduct of traditional 2-year bioassays.

D19. Jacobs (2005) compared the findings from short-term dose-ranging studies with the outcomes of 2-year rodent carcinogenicity studies for 60 pharmaceutical compounds in the CDER/FDA database. This evaluation considered liver, kidney, mammary, thyroid, adrenal, urinary bladder, lymph node/spleen, and lung. Contrary to the findings of Allen et al. (2004) (described in Section 3.1, above), short-term indicators such as hyperplasia, hypertrophy, increased organ weights, tissue degeneration or atrophy, and mineralisation were not reliable predictors of tumour outcome in the corresponding tissues in carcinogenicity bioassays. It was noted that some differences may be attributed to the different types of databases evaluated – many genotoxic and liver-toxic compounds are screened out in the pharmaceutical development process, there is greater variation in the rodent strains used for bioassays for pharmaceutical regulatory submissions than in NTP studies, and carcinogenicity bioassays for pharmaceuticals do not necessarily test the maximum tolerated dose.

NEG CARC

D20. Reddy et al. (2010), tested a 'whole animal negative predictivity' strategy, finding, in agreement with Jacobs (2005), that histopathological changes indicative of hyperplasia, cellular hypertrophy, and atypical cell foci were not reliable predictors of tumour outcome in the corresponding tissues. However, the complete absence of histopathological evidence of pre-neoplasia in all tissues in short-term toxicity studies was a reliable indicator for negative tumour outcome in a 2-year bioassay. In this study, 2-year rat bioassay data for 80 pharmaceuticals from commercial and Merck databases (30 carcinogens and 50 non-carcinogens) were compared with findings from corresponding 6- or 12-month toxicity studies.

D21. The 'whole animal negative' model specified the presence of pre-neoplasia (hyperplasia, cellular hypertrophy, and atypical cellular foci) in any single tissue (25 of the 30 carcinogens) as positive, and the absence of pre-neoplasia in all tissues (35 of the 50 non-carcinogens) as negative (sensitivity 83%, specificity 70%, negative predictive value 88%, positive predictive value 63%[1]). The 5 false negatives (i.e. negative from analysis of 6-month data but positive for tumours in 2-year rat bioassays) were all negative in genotoxicity assays and 2-year mouse carcinogenicity bioassays, and all produced tumours in rats based on proliferative or hormonal effects. The authors considered that the positive 2-year rat bioassay results for these 5 compounds were of questionable relevance to carcinogenicity in humans. They were all approved compounds currently marketed for non-life-threatening specifications and tumourigenicity was considered to be associated with rat-specific mechanisms.

D22. A larger project incorporating data from 13 companies was set up to further test the whole animal negative predictivity strategy proposed by Reddy and colleagues, using an expanded database maintained by the Pharmaceutical Research and Manufacturers of America (PhRMA) and including 182 pharmaceutical compounds (66 positive and 116 negative in 2-year rat carcinogenicity studies) (Sistare et al., 2011). In this study, negative outcome was specified as the absence of all of three criteria:

- genotoxicity
- any knowledge or significant evidence of hormonal perturbation activity
- evidence of histopathologic risk factors of rat neoplasia in all tissues examined in the corresponding chronic rat toxicity study conducted at similarly matching doses to those used in 2-year carcinogenicity studies.

[1] Sensitivity=TP/(TP+FN)X100, Specificity=TN/(TN+FP)X100, positive predictive value=TP/(TP+FP)X100, negative predictive value=TN/(TN+FN)X100 (TP=true positive, TN=true negative, FP=false positive, FN=false negative)

D23. This approach was termed 'NEG CARC' (Negative for Endocrine, Genotoxicity, and Chronic Study Associated Histopathological Risk Factors for Carcinogenicity).

D24. Immunosuppression was not included as a criterion on the basis that results in rat carcinogenicity tests do not reliably reflect human risk for this effect (Bugelski et al., 2010). It was noted there are likely to be significant differences between broad-based immunosuppressants and selective immune modulatory compounds that would be important to understand in helping to provide perspective for human risk assessment.

D25. *Genotoxicity* was assessed as any clear, single, positive genetic toxicology result in the good laboratory practice-compliant standard battery of assays that was not otherwise explained as an irrelevant finding.

D26. *Hormonal perturbation.* A weight of evidence approach was used, which included evidence of treatment-related microscopic and/or macroscopic changes in multiple endocrine tissues within a sex, measurements of changes in hormone levels, and knowledge of pharmacological mechanism of action (hormone receptor binding, alteration of hormone levels, alteration of activity of endogenous hormones).

D27. *Histopathology.* Positive findings were considered to be treatment-related hyperplasia, cellular hypertrophy, atypical cellular foci, or neoplasia in chronic studies (including multinucleated cells, basophilia, basophilic foci, cellular enlargement, cytomegaly, cellular swelling, cellular alteration, dysplasia, eosinophilic foci, karyomegaly, or tumour; excluding vaginal metaplasia and myocardial hypertrophy).

D28. The NEG CARC strategy identified 52 compounds as true positives (7 genotoxicity, 42 histopathology, 26 hormonal perturbation), 54 false positives (17 genotoxicity, 38 histopathology, 15 hormonal perturbation), 62 true negatives and 14 false negatives (sensitivity 79%, specificity 53%, negative predictive value 82%, positive predictive value 49% to predict rat carcinogenicity). Sensitivity was similar when considering endpoints at 6 or 12 months. As observed by Reddy et al. (2010), the sensitivity of microscopic findings to predict neoplasia in the 2-year rat study on an organ-by-organ basis was lower than on a whole-animal basis: for 9/42 true positives identified by histopathology, the tumour site in the carcinogenicity study did not match any of the positive tissues in the repeat-dose toxicity study (4 of these were considered to have hormonally linked mechanisms, 1 to be related to site of initial high exposure, and for 4 cases the mechanism was unknown).

D29. Eleven tissues (liver, thyroid, adrenals, ovaries, mammary gland, bone, pituitary, urinary bladder, kidneys, skin, stomach) served as sentinels in the 6/12-month studies for 90% of tumours in the 2-year studies. (The spectrum of positive

tissues for the histopathology false positives was noted to be similar). Nine sites accounted for over 80% of tumours (liver, thyroid, ovaries, testes, urinary bladder, skin, mammary gland, kidneys, adrenals). The authors suggested that tissues with the highest expected exposure after dosing or with high sensitivity to hormonal perturbations are thus most likely to be predictive of tumour risk to the rat. Many of the true positives were identified by early hormonal perturbation (often hormonal agents designed for this purpose). These were associated with ovarian granulosa cell, bone, mammary, testicular, pancreatic and/or thyroid tumours and all had earlier documented effects on hormones or hormonally regulated tissues in the rat in tissues related to the tumours seen in the lifetime bioassay. Development of such tumours in rats at sites distal to the primary drug target tissue was noted to be often due to rodent-specific mechanisms associated with chronic trophic hormonal stimulation at the target site for tumourigenesis that may or may not translate to humans.

D30. The human health relevance of positive 2-year rat bioassays for the 14 false-negative compounds was considered to be questionable and is discussed on a case-by-case basis. The overall conclusions were that the tumour signals were marginal, inconsistent across sexes, inconsistent across species and with a tendency to occur only at high doses. Ten of these compounds were marketed, 2 were not marketed for reasons unrelated to the rat carcinogenicity findings and 2 were still in development at the time of publication despite the positive rat carcinogenicity findings.

D31. An evaluation of data for 78 IARC Group 1 and 2A chemicals + 8 pharmaceuticals that had been withdrawn for cancer concerns was similarly carried out. Most of these (72) were positive for genotoxicity. Of the 14 non-genotoxic compounds, 10 would have been triggered for 2-year carcinogenicity testing by sub-chronic/chronic histopathology and/or known hormonal perturbation using the NEG CARC approach. Of the remaining 4 compounds, 3 were not carcinogenic in rats at doses that could be tolerated in 2-year studies (Group 1 – ethanol; Group 2A – 4-chloro-ortho-toluidine and tetrachloroethylene). Thus for ethanol (IARC Group I) the method would fail to predict the need to conduct a rat carcinogenicity study to identify a known human carcinogen, nevertheless the negative outcome of the rat study would have been correctly predicted. The other NEG CARC-negative compound (Group 1 – cyclosporine) was an immunosuppressant that would be expected to be negative in a 2-year rat assay but tumourigenic in humans.

D32. On the basis of this retrospective study, Sistare and colleagues proposed that a 2-year rat study is not necessary for compounds that are negative by the NEG CARC paradigm, and that human cancer risk assessment for such compounds can be carried out on the basis of a 6-month rat study + transgenic mouse study.

D33. Van der Laan et al. (2016) proposed that it would also be important to include

the category ‘pharmacological evidence’ as part of the NEG CARC approach. This hypothesis was tested in a detailed evaluation of a primary dataset of 298 pharmaceuticals, including 191 compounds from the ‘PhRMA’ database evaluated by Sistare and colleagues, 44 compounds from the CDER/FDA database, and 63 compounds from the Japanese Pharmaceutical Manufacturers Association (JPMA) database. Excluding 43 compounds that did not have a primary mammalian pharmacologic target (i.e. antivirals/antimicrobials), 255 compounds were categorised into 6 pharmaco-therapeutic areas (CNS, cardiovascular, respiratory, metabolic, hormonal, anti-inflammatory and immunomodulatory) plus ‘remaining’ compounds. Within these categories, 172 of the 255 compounds were sub-categorised into 45 pharmacological classes according to the primary drug target. Classification of tumourigenicity, based on the ‘NEG CARC’ criteria (genotoxicity + short-term histopathology + hormone perturbation) and on findings in 2-year rat bioassays was then correlated with pharmacological class.

D34. The aim was to identify pharmacological classes with a high proportion of positive class members. Ten of the 45 classes were ‘positive’ (contained > 50% compounds identified as rat carcinogens, see Table 9 from van der Laan et al., 2016, Annex 1), 17 classes were ‘negative’, and 18 classes had ‘mixed’ results. Not all compounds in each positive class were carcinogens, perhaps related to pharmacological, exposure, and replicability issues. Some compounds induced tumours considered to be unrelated to their pharmacology (e.g. induction of liver and/or thyroid tumours via induction of drug metabolising enzymes).

D35. Findings based on pharmacological class were discussed in the context of the NEG CARC prediction system for rat carcinogenicity, with particular reference to the 15 ‘false negatives’ present in the database investigated (11 from the PhRMA database evaluated by Sistare et al. (2011), 1 from the CDER/FDA database, and 3 from the JPMA database). For several compounds the inclusion of pharmacological class effects would have designated a true positive instead of false negative result, indicating that this could be a valuable additional criterion in a weight of evidence evaluation in cases where histopathology is negative at 6 months. The NEG CARC category ‘evidence for hormonal effects’ was proposed as too broad and better replaced with ‘oestrogenic/progestenic effects’. The spectrum associated with immunosuppressants was noted to be complex, with this class placed in the ‘mixed’ group (2 positive and 2 negative compounds in 2-year bioassay). The positives were an anti-TNFα compound that induced mammary gland tumours and systemic malignant lymphoma, and an immunosuppressant associated with granulocytic leukaemia in bone and interstitial cell tumours in testis. At least 1 of the 2-year bioassay negatives was positive in repeat dose toxicity studies (decreased thymus weight).

D36. Luijten et al. (2016) stated that the NEG CARC approach has also been tested in a retrospective analysis (*manuscript in preparation by* Woutersen et al.) of

around 200 ‘environmentally relevant chemicals’ using data from sub-chronic 90-day studies in rats, with findings in agreement with those of Sistare et al. (2011) that the absence of pre-neoplastic histological changes can accurately predict the lack of carcinogenicity of a non-genotoxic chemical.

Revision of ICH Guideline S1

D37. The International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH), in 1998, in its Guideline S1, proposed that carcinogenicity testing of small molecule pharmaceuticals for regulatory purposes be based on a 2-year test in one (rather than, historically, two) rodent species, supplemented with other data (a short- or medium-term *in vivo* rodent test or a second long-term carcinogenicity test in another rodent species) (ICH, 1998). Approaches using transgenic mice have subsequently been adopted (reviewed in G07 Part A), while the utility of other short-term study data is currently being evaluated.

D38. Ongoing revision of ICH S1 now aims to define situations where complete waiver of a 2-year bioassay would be justified (ICH, 2016a). A recent review of this process noted that the various available datasets that have been evaluated retrospectively have indicated that sufficient information should be available from pharmacology, genotoxicity and chronic toxicity data to conclude that a given pharmaceutical in certain cases presents a negligible risk or, conversely, a likely risk of human carcinogenicity without conducting a 2-year rat carcinogenicity study. Compounds could thus be listed in one of three main categories:

- Category 1 - highly likely to be tumourigenic in humans such that a product would be labelled accordingly and 2-year rat, 2-year mouse, or transgenic mouse carcinogenicity studies would not add value
- Category 2 - the available sets of pharmacologic and toxicologic data indicate that tumourigenic potential for humans is uncertain and rodent carcinogenicity studies are likely to add value to human risk assessment
- Category 3a - highly likely to be tumourigenic in rats but not in humans through prior established and well recognised mechanisms known to be human irrelevant, such that a 2-year rat study would not add value, or Category 3b - highly likely not to be tumourigenic in both rats and humans such that no 2-year rat study is needed. A 2-year or transgenic mouse study would be needed in most cases.

D39. A set of weight of evidence criteria has been developed to assign candidate compounds to these categories, including: knowledge of intended drug target and pathway pharmacology, secondary pharmacology, and drug target distribution in rats and humans; genetic toxicology study results; histopathologic evaluation of repeated dose rat toxicology studies; exposure margins in chronic rat toxicology studies;

metabolic profile; evidence of hormonal perturbation; immune suppression; special studies and endpoints (e.g. emerging technologies, new biomarkers..); results of non-rodent carcinogenicity study; transgenic mouse study.

D40. The ICH S1 revision process has been reviewed in the publication by Morton et al. (2014).

D41. The ICH is currently evaluating prospectively the reliability of this less-than-lifetime strategy through data generated by companies and will base their guidance on the outcome of this exercise (ICH, 2016a). Carcinogenicity assessment documents submitted by sponsors based on the weight of evidence factors will be evaluated before completion of 2-year bioassays, allowing regulatory agencies to assess how well the weight of evidence predicts the 2-year rat carcinogenicity study results. The 'prospective evaluation period' for this work began in 2013 and is currently expected to have gathered sufficient data to assess the viability of the weight of evidence approach by the end of 2017, with the final study report expected to be submitted at the end of 2019 (ICH, 2016b). The goal is to evaluate carcinogenicity assessment documents plus 2-year data for 50 compounds, at least 20 of which are in Category 3.

4 Integrated Approaches For The Identification And Risk Assessment Of Human-Relevant Carcinogens

D42. The studies reviewed in Section 3 used data from rodent carcinogenicity bioassays as the comparator, i.e. the approach taken has generally been to evaluate the effectiveness of short-term tests to predict the results of carcinogenicity assays in rodents rather than directly addressing the likelihood of carcinogenicity in exposed humans. In addition, many strategies focus on carcinogen hazard identification and may support labelling requirements, but do not address the potential spectrum of risk over a range of exposure levels. They are therefore less well suited to the assessment of levels of carcinogenic risk posed by chemical exposures at ambient levels present in the human environment.

D43. A key issue that is re-iterated by many commentators is the need to move to a strategy based on the identification of human-relevant carcinogens (Meek et al., 2003). For the incorporation of short-term tests into such a strategy, it is necessary to establish which short-term data are required to achieve this. This should be informed by consideration of the key events and modes of action of carcinogenicity (see COC discussion paper CC/2016/08). Rodent-specific modes of action would be excluded from the strategy as the aim is to identify and evaluate human-relevant carcinogens. A combined *in vitro* and *in vivo* approach may be developed, with an initial evaluation for *in vitro* signals that might indicate carcinogenic potential (e.g. genotoxicity tests, high-throughput screening) and subsequent confirmation of the relevance or otherwise in short-term *in vivo* tests. Some generic key events (e.g. cell

proliferation, immunosuppression) may be evaluated as short-term endpoints *in vivo*. Toxicogenomic techniques (e.g. transcriptomics) may also be applied to the evaluation of additional endpoints/biomarkers incorporated into sub-chronic toxicity studies, which may be able to highlight carcinogen class-specific signatures (*reviewed in* Doktorova et al., 2012). Toxicogenomic and high-throughput screening approaches are addressed in G07, part c and are not discussed in detail here.

4.1 Tiered and weight of evidence-based strategies to predict human carcinogenicity that incorporate parameters measured in sub-chronic toxicity studies

D44. Strategies and paradigms have been proposed that incorporate findings from short-term *in vivo* endpoints into human carcinogenicity risk assessments based on tiered and/or weight of evidence approaches.

D45. Cohen (2004, 2010a,b) has argued that the 2-year rodent bioassay is no longer necessary or appropriate for the evaluation of possible carcinogenic risk of chemicals to humans and that its use should be discontinued. An alternative model is presented that is based on shorter term tests, with an emphasis on mode of action and interpretation of the relevance to humans of findings in rodents. The premise is that increased carcinogenic risk occurs via: 1. increased net rate of DNA damage per cell division, occurring in pluripotential cell populations, and 2. increased number of DNA replications – i.e. increased cell proliferation (either by direct mitogenesis involving hormones or growth factors, or by cytotoxicity and regenerative proliferation) or decreased cell loss (by inhibition of apoptosis or cell differentiation). The model is represented as a tiered approach, incorporating a short-term screen for genotoxicity, immunosuppressive and oestrogenic activity using *in vitro* and *in vivo* tests, and the conduct of a 13-week assay using multiple doses to evaluate endpoints indicating toxicity/cell proliferation.

D46. The key events in this testing schedule involve precursor changes that can be identified in 13-week studies in rats and mice. The screening proposed has two phases: a general screen for any potential activity in any target tissue, and then a more detailed evaluation of the specific tissues identified as potential positives. The aim of this second stage is a careful mechanistic evaluation to identify the basis of the positive result, to determine whether the mode of action is relevant to humans and to define dose–response curves. This stage may eventually incorporate -omics methods.

D47. In this approach, *genotoxicity* (or DNA-reactivity) would be assessed by Ames assay and structure activity relationships, and by *in vivo* tests if necessary. Positives could then be evaluated for dose–response for DNA reactivity and for cell proliferative effects (which may occur at higher doses), to aid in extrapolation of the assessment to humans.

D48. *Oestrogenic activity* would be detected by *in vitro* assays and/or histologic assessment of typical oestrogen-affected tissues (e.g. breast, endometrium, cervix).

D49. *Immunosuppression* could be assessed by *in vitro* assays and/or in 13-week studies by histopathologic evaluation of immunologic (e.g. thymus, lymph nodes, spleen) tissues.

D50. *Toxicity and/or increased cell proliferation* could be demonstrated on the basis of histopathological examination, and possibly also using screens for DNA synthesis such as BrdU, PCNA or Ki-67 labelling index assays. Clinical chemistry and organ weight data from 13-week studies may be helpful.

D51. The question of how to evaluate the different tissues is noted to be a subject of debate. For example, examination of rodent tissues that do not have human counterparts (e.g. forestomach, Zymbal's gland, Harderian gland) may be of uncertain relevance, and species-specific tumours in rodents that have no analogue in humans (e.g. splenic mononuclear cell leukaemia in rats, mouse submucosal mesenchymal lesion of the urinary bladder) may have little predictive value for human tumourigenicity. The evaluation of rodent endocrine tissues for carcinogenic activity is proposed to be of limited predictive value for human cancer risk, except for the evaluation of oestrogenic activity, due to differences in kinetics, metabolism and dynamics of these tissues and feedback mechanisms between humans and rodents. Many of these tumours occur at high rates spontaneously in rodents, such as the thyroid, pituitary, and testicular Leydig cell tumours in rats. Rodents are resistant to some tumours at sites that are common in humans, such as glandular stomach, colon, prostate and pancreas. Conversely, liver, kidney, lower urinary tract and, to some extent, lung, tumours show some correlation between humans and rodents. The correlation is strongest for carcinogenesis induced by DNA-reactive compounds. Various modes of action have been identified, some of which are considered to be relevant and some irrelevant to humans.

D52. Luijten et al. (2016) also proposed a tiered test strategy for cancer hazard identification, incorporating existing knowledge, genotoxicity data and data from sub-chronic rat studies. This would include:

- Tier 1. Review of existing data (physico-chemical, toxicokinetic/dynamic, intended use, (quantitative) structure activity relationships
- Tier II. Genotoxicity tests *in vitro*
- Tier III. Genotoxicity tests *in vivo*
- Tier IV. Carcinogenicity.

A weight of evidence approach focussing on sub-chronic, repeat-dose toxicity data: histopathology (pre-neoplastic, proliferative or toxic lesions), organ weights, blood and urine chemistries and immunohistochemistry (e.g. Ki-67 as cell proliferation

marker), plus pharmacological mode of action in the case pharmaceuticals. The authors noted that this strategy was developed to allow rapid implementation and does not fully address existing needs for mode of action information.

4.2 Proposal for an IATA for non-genotoxic carcinogens (OECD)

D53. Jacobs et al. (2016) (for the OECD) proposed the development of an IATA (Integrated Approach to Testing and Assessment)[2] to evaluate the carcinogenic potential of chemicals that are negative in genotoxicity screens, noting that the potential for carcinogenicity via non-genotoxic mechanisms is often not evaluated, due to the testing approach recommended under many regulatory frameworks.

D54. The selection of elements in an IATA can be based on an adverse outcome pathway concept incorporating biological changes, or key events, at the cellular, tissue, organ and organism levels that occur in response to molecular initiating events and leading to an adverse outcome. The relationships between molecular initiating events, key events and adverse outcomes are described in key event relationships. An IATA can also be developed empirically, containing elements other than those informed by the adverse outcome pathway, such as intended use and exposure, toxicokinetics and toxicodynamics.

D55. The proposed IATA comprises a structured information level framework with five levels of test information:

- **Level 0** incorporates available literature and *in silico* mode of action review information.
- **Level 1** (sub-cellular) and **Level 2** (cellular) tests evaluate mode of action groups *in vitro*, looking for molecular initiating events and early/initial key events. It is noted that widely accepted Level 1/2 test methods currently exist only for endocrine molecular initiating events (e.g. oestrogen receptor binding and transactivation, steroidogenesis). The ToxCast programme is cited as potentially useful for Level 1/2 tests, and also toxicogenomic approaches using *in vitro* test systems that group chemicals according to specific modes of action. A wide range of modes of action should be tested. Quantitative information such as dose–response relationships and points of departure will be required in order to be able to predict whether one key event would trigger the next key event.

[2] The OECD working definition of an IATA is: ‘a structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted testing and therefore optimising and potentially reducing the number of tests that need to be conducted’ (OECD, 2015).

- **Level 3** (multicellular tissue/organ) aims to identify cytoskeletal, tissue and organ changes and angiogenesis. It includes *in vitro* tests such as cell transformation assays and 3D cell models, *ex vivo* organ studies, *in vivo* data such as histopathology from repeat dose studies, and 'organ-on-a-chip' technologies. Level 3 *in vivo* information may not be needed if Level 2 and Level 3 *in vitro* data are sufficient to meet regulatory requirements based on molecular initiating event and weight of evidence information.
- **Level 4** (organism) includes transgenic rodent assays, 2-year rodent carcinogenicity bioassays and chronic toxicity studies. The aim is to minimise the necessity for Level 4 data, in line with the principles of 'Toxicity Testing in the 21st Century' (described in G07, part c). This information may be required to gain insight into adverse effect levels, dose–response curves and tumour types/species affected.

D56. Assays or diagnostic tools may overlap two levels (an example given is cell transformation assays, which may belong to Levels 2 and 3). Quantitative and qualitative adverse outcome pathway/mode of action elements are required, based on the steps of the carcinogenic process. All modes of action should be tested for (as blocks of tests), and negative results for one block should not exclude all other modes of action. It should also be noted that mechanisms are not always related to adversity, with early molecular initiating events/key events not always leading to downstream adverse outcomes. IATA-based decisions may be made when several interconnected mechanisms are affected adversely (for example, all of the three hallmarks – oxidative stress, cell death, immune system evasion). Level 1, 2 and 3 assays require validation such that definitive decisions including the derivation of acceptable exposure levels can be made.

5 Summary

D57. For several decades, the standard method used to evaluate the carcinogenic potential of chemicals has been a genotoxicity test battery plus extrapolation from the results of high-dose 2-year rodent bioassay tests to low-dose exposures in humans. Key drawbacks of this approach are the high number of false positive results obtained and the question of relevance to human cancer risk, due to issues of species specificity, mode of action, and dose.

D58. Alternative strategies to the 2-year rodent bioassay are being developed that incorporate short-term data into carcinogenicity evaluations, based on tiered approaches and/or weight of evidence evaluations. Some of these approaches are likely to be feasible in the short term whilst others are more exploratory and it is not yet clear whether they will be feasible for risk assessment purposes. They vary depending on the type of compound being evaluated and the purpose of the evaluation. For use in application to the risk assessment of chemicals present in the

environment, new systems for carcinogenicity evaluation would ideally have the potential to produce organ-specific, dose-dependent information relevant to humans.

D59. Retrospective evaluations of existing databases have shown some utility of short-term *in vivo* test data to predict the outcomes of 2-year rodent bioassays, but with the development of further short-term endpoints necessary. A negative-predictive approach (the absence of short-term histopathologic risk factors in multiple tissues) has shown utility for screening out non-carcinogens, particularly in the evaluation of pharmaceuticals. A strategy for evaluation of pharmaceuticals using a weight of evidence approach is being tested prospectively by the ICH to define situations where a waiver of the requirement for a 2-year rodent carcinogenicity bioassay can be granted.

D60. New tiered/integrated strategies are being developed using a mode of action-based approach incorporating modes of action that are of relevance to humans but not those that are rodent specific. The question of which key events/modes of action should be evaluated is a developing area. A combined *in vitro*/*in vivo* approach may be developed, looking for any *in vitro* signals that might indicate carcinogenic potential (e.g. in high-throughput screening) and then confirming relevance or otherwise in short-term *in vivo* tests.

D61. An IATA for the evaluation of carcinogenic risks posed to humans by non-genotoxic chemicals is in development (OECD). The goal is for a strategy without animal testing, based on tests for key events and key events relationships, as this knowledge base expands.

6 COC conclusions on alternative testing strategies for carcinogens incorporating results from short-term tests

D62. Use of the 2-year rodent bioassay to evaluate the carcinogenicity of the vast numbers of untested chemicals that are currently marketed is not practical and alternative methods are required for this purpose. Genotoxicity tests can detect many, but not all, genotoxic carcinogens, and cannot detect non-genotoxic carcinogens.

D63. The development of alternative approaches for the identification and characterisation of chemical carcinogens is a rapidly evolving field. Currently available data do not give a clear indication of the direction of progress in replacement. Some of the approaches that have been used have conceptual problems and there are currently no methods that are generally accepted in replacement of animal carcinogenicity studies.

D64. One approach that is being developed is the use experimental data from shorter-term tests, which may be incorporated into evaluations based on tiered

approaches or weight of evidence evaluations. Retrospective studies have indicated some utility of short-term (e.g. 3- or 6-month) *in vivo* test data to predict the outcomes of 2-year rodent carcinogenicity bioassays but further short-term endpoints are required.

D65. While short-term studies in vivo may be used as part of the weight of evidence to provide an indication that a chemical is potentially carcinogenic, they would be unlikely to provide a basis for estimation of tumour risk.

D66. In some cases the positive results observed in short-term tests in vivo have little biological plausibility. The signals indicating potential carcinogenicity are sometimes identified in different tissues to those in which tumours are identified in longer term studies, in addition to which the signals in the short-term assays may be hypertrophy or hyperplasia, which are not of themselves pre-neoplastic effects. This question of biological plausibility adds weight to the view that these short-term assays cannot be used as a replacement for the 2-year bioassay for the identification of rodent carcinogens.

D67. Negative predictive approaches, which incorporate negative outcomes for genotoxicity and short-term histopathologic risk factors in multiple tissues, are of interest but have associated problems, including the fact that human metabolism may not be suitably accounted for in genotoxicity tests. These approaches often have very high false-positives rates whilst not entirely excluding false-negative results when compared with the outcomes of 2-year bioassays. The challenge in the use of these methods is to identify all events that are crucial, while avoiding over-predictivity that could lead to the use of over-precaution.

D68. There is a limit to how far animal tests can be developed to predict human cancer risk, due to issues of species specificity. The emphasis should now be moved away from the development of further rodent studies. Future approaches should take into account human-relevant modes of action and alternative strategies should focus on predicting potential human carcinogenicity rather than rodent carcinogenicity.

D69. There are differences in approach between testing of pharmaceuticals compared to other chemicals. In the 2-year bioassay, the maximum dose tested for pharmaceuticals is generally equivalent to a large multiple of human exposure, while other chemical sectors tend to use the maximum tolerated dose. Pharmaceuticals are also in themselves associated with a pharmacological effect in humans, whereas other chemicals either have only a technical purpose in the media they are in (e.g. food additives), or are tested to ensure they do not show adverse effects in non-target species (e.g. pesticides). It is important to recognise that different approaches may be required and that alternatives may not address the requirements of all the different sectors. Nevertheless, it is important to maintain the collaborative approach across the sectors.

D70. Overall, it is important for alternative means of assessing health risks from chemicals to be developed and for good interaction between different sectors to continue. The challenge for such alternative strategies is to avoid missing crucial adverse effects while not over-predicting issues of concern.

References

Adeleye Y, Andersen M, Clewell R et al. (2015). Implementing Toxicity Testing in the 21st Century (TT21C): Making safety decisions using toxicity pathways, and progress in a prototype risk assessment. Toxicology, 332, 102-111.

Alden CL, Lynn A, Bourdeau A et al. (2011). A critical review of the effectiveness of rodent pharmaceutical carcinogenesis testing in predicting for human risk. Vet Pathol,
48, 772-284.

Allen DG, Pearse G, Haseman JK and Maronpot RR (2004). Prediction of rodent carcinogenesis: an evaluation of prechronic liver lesions as forecasters of liver tumors in NTP carcinogenicity studies. Toxicol Pathol, 32, 393-401.

Benigni R. Evaluation of the toxicity forecasting capability of EPA's ToxCast Phase I data: can ToxCast in vitro assays predict carcinogenicity? J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 2013;31(3):201-12

Benigni R. Predicting the carcinogenicity of chemicals with alternative approaches: recent advances. Expert Opin Drug Metab Toxicol. 2014 Sep;10(9):1199-208.

Bhattacharya S, Zhang Q, Carmichael PL et al. (2011).Toxicity Testing in the 21st Century: Defining New risk Assessment Approaches Based on Perturbation of Intracellular Toxicity Pathways. PLoS One, 6, e20887.

Boobis AR, Cohen SM, Dellarco V et al. (2006). IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol, 36, 781-792.

Boobis AR, Cohen SM, Doerrer NG et al. (2009). A data-based assessment of alternative strategies for identification of potential human cancer hazards. Toxicol Pathol, 37, 714-732.

Bugelski PJ, Volk A, Walker MR et al. (2010). Critical review of preclinical approaches to evaluate the potential of immunosuppressive drugs to influence human neoplasia. Int J Toxicol, 29, 435-466.

Cohen SM (2004). Human carcinogenic risk evaluation: an alternative approach to the two-year rodent bioassay. Toxicol Sci, 80, 225-229.

Cohen SM (2010a). An enhanced 13-week bioassay: an alternative to the 2-year bioassay to screen for human carcinogenesis. Exp Toxicol Pathol, 62, 497-501.

Cohen SM (2010b). Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay is no longer necessary. Toxicol Pathol, 38, 487-501.

Cox LA, Popken DA, Kaplan AM et al. (2016). How well can *in vitro* data predict *in vivo* effects of chemicals? Rodent carcinogenicity as a case study. Regulatory Toxicology and Pharmacology, 77, 54-64.

Doktorova TY, Pauwels M, Vinken M et al. (2012). Opportunities for an alternative integrating testing strategy for carcinogen hazard identification. Crit Rev Toxicol, 42, 91-106.

Hanahan D and Weinberg RA (2000). The hallmarks of cancer. Cell, 100, 57-70.

Hanahan D and Weinberg RA (2011). Hallmarks of cancer: The next generation. Cell, 144, 646-674.

ICH (1998). ICH Topic S1B. Carcinogenicity: Testing for carcinogenicity of pharmaceuticals, March 1998.

ICH (2016a). ICH Guideline S1. Regulatory notice on changes to core guideline on rodent carcinogenicity testing of pharmaceuticals. 29 February 2016. Accessed at: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2012/12/WC500136405.pdf, 07/06/16.

ICH (2016b). The ICHS1 Regulatory Testing Paradigm of Carcinogenicity in rats – Status Report. 2 March 2016. Accessed at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S1/Final_Status_Report_ICH_S1_Jacksonville_2March2016.pdf, 07/06/16.

Jacobs A (2005). Predication of 2-year carcinogenicity study results for pharmaceutical products: how are we doing? Toxicol Sci, 88, 18-23.

Jacobs MN, Colacci A, Louekari K et al. (2016). International regulatory needs for development of an IATA for non-genotoxic carcinogenic chemical substances. ALTEX Online first, Published April 27, 2016 http://dx.doi.org/10.14573/altex.1601201

Julien E, Boobis AR, Olin SS (2009). The Key Events Dose-Response Framework: a cross-disciplinary mode-of-action based approach to examining dose-response and thresholds. Crit Rev Food Sci Nutr, 49, 718-728.

Kleinstreuer NC, Dix DJ, Houck KA et al. (2013). *In Vitro* Perturbations of Targets in Cancer Hallmark Processes Predict Rodent Chemical Carcinogenesis. Toxicological Sciences, 131, 40-55.

Long, GG (2010). Commentary on “Evaluation of possible carcinogenic risk to humans based on liver tumors in rodent assays: the two-year bioassay I no longer necessary.” Toxicol Pathol, 38, 502-505.

Luijten M, Olthof ED, Hakkert BC et al. (2016). An integrated test strategy for cancer hazard identification. Crit Rev Toxicol, published online 03 May 2016, http://dx.doi.org/10.3109/10408444.2016.1171294

Meek ME, Bucher JR, Cohen SM et al. (2003). A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol., 2003, 33, 591-653.

Morton D, Sistare FD, Nambiar PR et al. (2014). Regulatory Forum commentary: alternative mouse models for future cancer risk assessment. Toxicol Pathol, 42, 799-806.

NRC (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. N. A. Press.

OECD (2015). Report of the workshop on a framework for the development and use of integrated approaches to testing and assessment. Series on Testing and Assessment No. 215. Accessed at: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2015)22&doclanguage=en 07/06/16.

Reddy MV, Sistare FD, Christensen JS et al. (2010). An evaluation of chronic 6- and 12-month rat toxicology studies as predictors of 2-year tumor outcome. Vet Pathol, 47, 614-629.

Schwetz B and Gaylor D (1998). Alternative tests: carcinogenesis as an example. U.S. Food and Drug Administration/National Center for Toxicological Research, Jefferson, Arkansas.

Sistare FD, Morton D, Alden C et al. (2011). An analysis of pharmaceutical experience with decades of rat carcinogenicity testing: support for a proposal to modify current regulatory guidelines. Toxicol Pathol, 39, 716-744.

U.S. EPA (2013). A retrospective analysis of the immunotoxicity study (OCSPP test guideline No. 870.7800). February, 2013. Accessed at:http://www.epa.gov/pesticide-registration/retrospective-analysis-immunotoxicity-study 07/06/16.

van der Laan JW, Kasper P, Silva Lima B et al. (2016). Critical analysis of carcinogenicity study outcomes. Relationship with pharmacological properties. Crit Rev Toxicol, Published online 26 Apr 2016, http://dx.doi.org/10.3109/10408444.2016.1163664