Criteria for a Recommended Standard:
Occupational Exposure to Diacetyl and 2,3-Pentanedione

Response to Peer Review Comments

The attached National Institute for Occupational Safety and Health (NIOSH) Response to Diacetyl and 2,3-Pentanedione Peer Review Comments document provides a complete listing of all peer reviewer comments received and the NIOSH response to each comment during both peer review processes. Peer review comments are provided in their entirety within this document and are organized by topic area. This format allows ease of use for the reader by providing all perspectives of a similar topic area together. In some cases, peer reviewers made similar comments. In these cases, NIOSH responded to the comment the first time it is listed and then referenced the same response for subsequent comments. In accordance with NIOSH policy, attribution of specific peer reviewer comments is blinded in this document. Information from the public meeting, external review draft document for public comment, and public and peer review comment submissions from the 2011 public comment period are available on the NIOSH Diacetyl and 2,3-Pentanedione Criteria Document Docket page: http://www.cdc.gov/niosh/docket/archive/docket245.html. Information from the re-review of Chapter 6 and new section of Chapter 8, including the external review draft chapters and the public comment submissions are available on the Regulations.gov website at https://www.regulations.gov/document?D=CDC-2013-0021-0001.

Abbreviations

BM	Benchmark dose
BMDL	Lower confidence limit of the benchmark dose
BMR	Benchmark response value
CFD/PBPK	Computational fluid dynamics/physiologically-based pharmacokinetic
EPA	Environmental Protection Agency
FEV_1	Forced expiratory volume in 1 second
FVC	Forced vital capacity
GHS	Globally Harmonized System of Classification and Labeling of Chemicals
HEC	Human equivalent concentration
HRX	High-risk exposure
hr	Hour
LLofN	Lower limit of normal
LOAEL	Lowest observed adverse effect level
LOD	Limit of detection
N	Number
NFPA	National Fire Protection Association
NIOSH	National Institute for Occupational Safety and Health
NOAEL	No observed adverse effect level
NTP	National Toxicology Program
OSH	Occupational safety and health
OSHA	Occupational Health and Safety Administration
PEL	Permissible exposure limit
PFT	Pulmonary function test
POD	Point of departure
Ppb	Parts per billion
Ppm	Parts per million
REL	Recommended exposure limit
STEL	Short-term exposure limit
TERA	Toxicology Excellence for Risk Assessment

TWA	Time-weighted average
UF	Uncertainty factor
VE	Volume estimate
yr	Year

Tracking Number	Reviewer's comment	Response
5001	First I want to congratulate the team who developed this document and all of the background research over the past decade for producing landmark research findings and presenting a credible approach to handling a new serious hazard in a very complex context. Neither the magnitude nor the scope of this daunting task is lost on this reader. Although this is not my area of day-to-day work the review appears to be thorough and to cover the available literature completely as best I can discern.	No response needed
5002	That said, I have one major concern about the presentation, and one perhaps larger concern about the strategy ultimately used for deriving the REL. Re the presentation: The report in its present form incorporates vast amounts of information about and derived from the various exposure settings of concern, from the manufacture of popcorn flavoring materiEs, through the manufacturing of other flavors, through the many other sectors of the food and beverage industry (where these flavors or naturally occurring sources of diacetyl and related chemicals may occur) all the way to end users of popcorn and other naturally or artificially flavored products. This is daunting, especially to a reader not very familiar with this highly complex industry. For the complex data-synthesis to be communicated effectively, there needs to be a *single, clear, user-friendly and consistent "road-map"* for matching the exposure and health-effects data with the sector, thereby allowing the reader to comfortably integrate the very complex differences in exposures, numbers of people exposed, available IH and biomedical data. At the end of two readings I was left with the impression that while the vast bulk of the credible human data derive from microwavable popcorn flavor manufacturing, the document incorporates data and makes inferences—especially regarding the scope of concern and application of an REL and associated recommendations—on a vastly broader and more complicated worker population. It would be well worth the effort in my view to have an introductory "road-map"	Chapter 1 provides a general overview of industries that have employees potentially exposed to diacetyl and 2,3-pentanedione. We have also enhanced the descriptions of exposure and health effect studies across industries in Chapter 2 and Chapter 3. We believe these three chapters provide a comprehensive review on the existing literature. As a result, we have not added a roadmap as suggested. The peer reviewer also suggests that diacetyl exposures and resulting health effects outside the microwave popcorn environment could be different. We respectfully disagree as shown by health effects in a variety of workplace settings such as cookie manufacturers, coffee roasters, and flavor manufacturing [Kanwal et al. 2006; Martyny et al. 2008; NIOSH 2003a, b, 2004a, b, 2006, 2007, 2008a, b, 2009, 2011b]. To investigate the concern of the non-generalizability of the microwave high exposure (mixers') experience, we have conducted additional analyses in which workers who were ever mixers were excluded; the estimates for the diacetyl effect were larger in this new analysis presented in Chapter 5.

	chapter delineating the sectors of the industry, outlining what information about exposure and health is available from each (and over what time course, as many of these are very mature industries I assume and the absence of earlier documented effects may be relevant), then sticking religiously with that outline throughout the report rather than parsing each of the subsequent sections differently. This recommendation is not just for cosmetic purposes: as will be clearer from comments below, I had enormous difficulty teasing out what credible evidence there is regarding these exposures *outside* of micro-popcorn manufacturing, as a consequence of which some very substantial reservations about the document's bottom line.	
5003	My more substantive issue relates to the sequence of inferences which underlie the proposed REL. I for one find the core conclusion from the human (largely HHE) data—that uncontrolled exposures during the mixing and packaging of micro-popcorn flavorings *cause* bronchiolitis obliterans—highly compelling and unworthy of further debate. Arguments raised by naysayers regarding the absence of biopsies in many cases, the use of case-series rather than formal epidemiologic testing methods, and "variability in the presentation of the cases" are just, frankly, gorilla dust.	Thank you. Note, however, that the recommended exposure limit (REL) is not based on bronchiolitis obliterans, but rather on lung function.
5004	On the other hand, I am less convinced by many of the further inferences drawn, starting with the most troubling which is the specificity of the relationship to diacetyl. Although it is far and away the most likely culprit, the data as presented do not extend to a serious consideration that either another closely correlated exposure in micro-popcorn, or some other aspect unique to this exposure setting, may be important. Likewise, I am not convinced by the causal connection to the other reported health outcomes such as restrictive lung disease or asthma. Moreover, even assuming diacetyl is the singular cause of BO and possibly other respiratory effects in the micro-popcorn setting, the short-term effects of exposures at *lower doses* or *long-term low-level settings* remains, in my view, quite speculative: absent a very clear understanding of the contribution of exposures that occurred before quantitative IH was	We respectfully disagree with the peer reviewer comment and provide a solid case for causation between diacetyl and health outcomes in Chapter 3. As presented in Chapter 3, health effects from diacetyl have been observed in industries outside microwave popcorn. See Table 3.1.

	conducted at the personal level, all “extrapolations”, either in time or dose, are fraught with assumptions that have only modest support from the human epidemiologic data as presented, or as best one can decipher it given the presentation issue raised above. Put another way, I am not convinced that low-level exposure over long periods of exposure *outside* of micro-popcorn flavoring has been proved to cause significant lung disease in workers (although there are suggestions), posing a substantial dilemma for control.	
5005	Moreover, although I am not a toxicologist so take this with the appropriate grain of salt, I find the risk assessment based on animal models equally uncompelling as a firm basis for an REL—neither is the target effect from the sub-chronic models obviously relevant for the BO end-point of interest in humans, nor would structure-activity inferences based on diacetyl lead one to anticipate the occurrence of BO or any other well-defined respiratory effects at the levels of concern, i.e., ppb levels.	Computational fluid dynamic modeling suggests that a larger fraction of an inhaled dose of diacetyl will reach the deep lung in humans than in rodents; therefore, it is not surprising that the most sensitive respiratory target region in mice may differ from that observed in humans. The toxicologically-based risk assessment is predicated on the assumption that the tissue dose of diacetyl that causes respiratory toxicity in mice may also cause respiratory toxicity in humans, even if that toxicity is observed in a different region of the respiratory tract.
5006	All these issues conspire to become a significant problem for establishing an REL, based largely by extrapolation from cases likely exposed at levels >1000-fold higher for at least some period of time historically and for the most part in one specific exposure context—micro-popcorn flavor manufacture. Since the REL would impact a very broad array of exposure contexts, and largely impact workers who under any circumstance would have far lower exposures to diacetyl and other chemicals involved in food flavoring, the justification is incomplete.	Analyses that excluded the highest exposed workers (mixers) produced the same findings as in the full population. In the analyses without mixers, the highest exposed workers had diacetyl levels of approximately 0.5 ppm, which is 100-fold higher than the proposed REL. The REL corresponds to a 45-year (yr) work life. Moreover, as stated in the document, a substantial proportion (17%) of the workers had career-average exposure levels below 0.01 ppm, which is only 2-fold above the REL.

5007	Of all of the immediate solutions to the problem, the one that would be most helpful would be—following my first suggestion above—a separate review of human data derived from each of the non-popcorn settings separately, or, if NIOSH prefers to approach the problem using the unifying assumption that the exposure to diacetyl is the only issue (hence the setting doesn't matter, only dose and duration) then at least it would be helpful to review the evidence that remains once the cases occurring in the high-exposure micro-popcorn setting are removed. Is there a corpus of evidence that loss of lung function--with or without the pattern of BO—occurs exclusively in lower dose settings in a dose-response pattern? And if so, what do those slopes look like (with the ultra-high exposure subjects removed)? From my perspective this is the kind of epidemiologic analysis which, even if confidence intervals are wide, would provide a secure basis for an REL in the range proposed. Frankly it is very possible the data are there, only too hard to ferret out for this reader given the presentation.	The human data in Chapter 3 are organized into sections on microwave popcorn, on flavoring manufacture, on diacetyl manufacture, and on other food production case reports. In response to the reviewer's concern, the exposure estimation for flavoring manufacture and diacetyl manufacture is inadequate for determining quantitative exposure-response relations. In the case of flavoring manufacture, batch operations preclude obtaining representative diacetyl exposures cross-sectionally. In the case of diacetyl manufacture, the historical estimates were limited. However, in microwave popcorn manufacture, cases consistent with bronchiolitis obliterans occurred in persons with orders of magnitude lower exposure than mixers in the sentinel Missouri facility (Facility G) and with short latency. For example, mixers developed lung disease in Facility K in which personal exposures averaged 0.02 ppm, although peak exposures as high as 80 ppm were demonstrated with manual pouring of liquid flavorings into heated oil. Packaging area workers had higher prevalences of airways obstruction in plants without isolated mixing operations with average personal exposures in the range of 0.5–0.6 ppm. Thus, exclusion of sentinel facility workers or mixers in analyses still results in risk. These data are already summarized, excluding Facility G workers and mixers in any facility in the draft criteria document in section 3.1.2.4, relying on Kanwal et al. [2006]. The data on longitudinal loss of lung function in relation to quantitative exposure is limited because serial NIOSH testing was unavailable in the microwave popcorn industry apart from in the sentinel facility (Facility G). The serial spirometry data over one year generated in a company-sponsored study [Lockey et al. 2009] addresses some of the reviewer's concerns: Even after institution of a powered air purifying respirator requirement for mixers, mixers without previous high exposures had a statistically significant 5.7-fold increased risk of airways obstruction compared to never-mixers (section 3.1.2.5), despite anticipated lowering of diacetyl exposure by 25-fold. A cumulative exposure of 0.8 ppm-years or more was associated with airways obstruction. They found no association between excessive longitudinal change in FEV_1 over one year in a

		subset of workers with three measurements with current average exposures <0.05 ppm and ≥0.05 ppm diacetyl (uncorrected). However, these data could be interpreted as consistent with 0.05 ppm being above a threshold for excessive pulmonary function decline, having insufficient power, or reflective of longitudinal FEV_1 decline being less related to current exposure than to previous exposure, as was found for the sentinel facility workers after interventions lowered diacetyl measurements to below detection limits for most workers [Kanwal et al. 2006]. Longer term follow-up of the Lockey et al. [2009] cohort over 6 years may shed light on these concerns for pulmonary function declines in relation to exposure. These analyses are underway. The reviewer suggested that it was possible that we had quantitative diacetyl data to address his concern about supporting an REL in the parts per billion (ppb) range from serial pulmonary functions, but that he could not ferret it out. The analytical limit of quantification in the low ppb range was not available for NIOSH field investigations in the microwave popcorn industry. In addition, comparatively short-term serial spirometry data have variability and result in imprecise slopes. We can demonstrate consistent lung disease outcomes in a variety of industries and companies, but we do not have adequate exposure characterization apart from microwave popcorn to address his question without modeling of risk that extrapolates beyond the data in hand.
5008	I believe you have presented the relevant literature completely and fairly.	No response needed
5009	No.	No response needed
5010	I believe the overall approach to control is reasonable, although this is less my area of expertise.	No response needed
5011	This is the subject of my initial review and answer to the supplementary questions. I fear that the proposed REL may on the one hand be drawn from a narrow subset of exposed groups—largely the micro-popcorn workers—yet is being applied to a very broad array of worker groups for whom little or no evidence of risk has been demonstrated.	The REL is based on microwave popcorn workers because representative exposure characterization in flavoring manufacturing workers is not available because of the batch processes involved in production. Certainly flavoring manufacturing workers have risk, and case reports exist of workers exposed to flavorings in other food production industries, such as flavored

		coffee production and candy manufacture. Cases of constrictive bronchiolitis were not recognized in microwave popcorn and flavoring manufacture for many years, so the absence of recognized cases in the many industries that use flavorings is not reassuring. The extrapolation of the microwave popcorn data to low level exposure is another issue. Excluding the job categories with highest flavoring exposures, such as mixers, did not result in the disappearance of cases in microwave popcorn plants. Indeed, in Facility G, packaging workers ended up on lung transplant lists within months of first exposure, despite having average exposures an order of magnitude lower than mixers and four times higher than quality control workers (with greater prevalence of abnormality). No changes have been made to the document in response to this comment.
5012	As I have tried to emphasize in the answers to additional questions, this question is unanswerable from the data because there is no compelling theory or evidence upon which to rely. While the 45 year duration is more conservative, almost all of the cases recognized have occurred from high-dose, short-exposure. The animal model sheds no further light on this, and there is not even a strong hypothesis regarding mechanism of action which might help. For this reason I propose you present both versions and select the one you believe is appropriate based on how conservative you choose to be, rather than on convention.	We have repeated analyses removing workers who had ever been mixers and see slightly *stronger* estimates of exposure response. This provides evidence that the effects are not largely arising from short-term high exposures. See also response to comment 5092.
5013	No.	No response needed.
5014	Comment: I agree with the exclusion of plant G from the risk analysis. Re the choices for K and L, the overarching problem is that the data are scant. Either choice has merit but neither choice solves the underlying dilemma. If the two are combined there should be a model term for plant fixed effects, which likely would nullify the "gain"…. I always favor doing both when in doubt as a sensitivity test.	The criteria document presents a thorough rationale for basing the risk assessment on Plant G and explicitly not on Plants K and L. Had we chosen Plants K and L, the proposed REL would be substantially lower than 0.005 ppm.

5015	I prefer the narrower (more restrictive) definition, although NOT because I imagine any other outcome than BO to be interesting—it's a matter of more vs less sensitive criteria (hence also less vs more specific) for BO. As always it's a trade-off and as always the best way to explore the trade-off is to show both. Do keep in mind, however, that many subjects will meet the less restrictive criteria who have no disease (eg obesity) and a range of diseases un likely related to exposure.	Agreed. Our current REL is based on the less restrictive definition in the case of the BMD analysis and on the more restrictive definition for the modeling of incidence and estimating excess lifetime risk. Both give similar results. The variability in the pulmonary function is evident in our statistical models. With the less restrictive definition, more "cases" will be judged to be not attributable to the exposure. The more restrictive definition was used for rating model and excess lifetime risk. The excess lifetime risk has now been calculated for a third definition of case: forced expiratory volume in one second/forced vital capacity<lower limit of normal (FEV_1/FVC<LLofN) (called "definition 2" in the revised document; the most restrictive definition is now "3"). Lung function decline is what is estimated in the longitudinal analysis (Table 5.11). The BMD procedure presented here is only for continuous outcomes (FEV_1,..) not dichotomized. See also response to comment 5086.
5016	I agree with the general idea of avoiding the use of data in the risk analyses which could strongly bias the result upward, in this case any data that would substantially underestimate exposure. That said, my general choice is to present every analysis both ways and make your case as one would with a sensitivity analysis.	The problem is how to deal with date of onset when no symptoms beginning after exposure are reported because modeling of incidence rate requires a date of onset. If workers already qualified as a case based on pulmonary function test (PFT) at their first survey but reported no symptoms beginning after first exposure, they had to be excluded.
5017	The model is not the problem—paucity of data is the problem, especially paucity of "cases" and non-cases observed as a consequence of long-term longitudinal exposure. As a consequence it is very hard to draw any strong inference without heavy parametric assumptions about the time-dose-effect relationship. I propose as with all the answers above that each choice be shown and the relative merits of each described (more or less as you have done, albeit please tighten the language!) but once again you can't in the end expect models to solve the problem when there is not enough relevant observation—the truth is there is almost no basis other than by applying wisdom form other populations and toxins on which to infer the effect of long term low dose exposure, and "better" models don't help.	See response to comment 5100.

5018	The answer to this is he same as the above. Its not the choice of metric that limits inference but the paucity of the observations you need to address which pattern of exposure is most toxic. That said, there is much for taking a very "agnostic" view such as defining an average exposure level beneath which no cases seem to have occurred and deal with uncertainty the old-fashioned way, by use of protective factors.	The draft was revised to better explain the interpretation of susceptibility, although there is no available guidance on the issue of "average" exposure and its likely connection to the susceptibility phenomenon. If susceptibility is decreasing with duration, and therefore cumulative exposure becoming less important with increasing duration, it follows that average exposure (defined as cumulative exposure divided by duration) could be a reasonable predictor of risk in this strange situation. If susceptibility converges to some constant level, average exposure would no longer be predictive, especially extrapolating over 10 or 45 years. This is not primarily a data paucity issue, although with much more data, one could more reliably model the unexpected relationship involving duration. The importance of peak exposures is in doubt because removing workers who were ever mixers increases, not decreases, the exposure-response slope. See also responses to comments 5135 and 5132.
5019	While I agree the data suggest some workers were "immune" from effect, hence the dose-response in those susceptible is undoubtedly higher than for the population as a whole, absent any marker or even really insight into who susceptibles may be, I frown on the use of models to capture this quantitatively. My preference would be to add safety factors as above to take into account yet another level of uncertainty.	Applying safety factors requires some estimate of a lowest observed adverse effect level (LOAEL) or no observed adverse effect level (NOAEL) for the "low risk" population, which is not available from the regression models. The importance of peak exposures is in doubt because removing workers who were ever mixers increases, not decreases, the exposure-response slope. Also, the dose-rate analyses that observe a better fit with concentration raised to a power < 1.0 argue against a role for peak exposures. The existence of variable susceptibility is indeed speculative, as is any hypothesis-generation in the face of unexpected findings, in this case the divergent associations of incidence with duration vs. cumulative exposure. The high risk exposure (HRX) term (now called "shortdur(DA)") is actually a product of a duration factor and average exposure (or its square), rather than cumulative exposure. The term is applied to the entire population precisely because the susceptibility composition is entirely unknown. This term was constructed ad hoc based on a simple assumption of declining susceptibility and empirical model fit, in an attempt to describe the observed incidence rate. If "group 2" refers to high risk workers, they are not identifiable except in relation to time of onset, and

		there is no reason to assume that we are dealing with two discrete groups—it could be a continuum and could even be one group whose susceptibility changes over time with exposure (e.g., by some immune or other adaptation). Mixture modeling would be problematic if finite fixed groups cannot be assumed, and statistical power would also become an issue with estimation of additional parameters. The document has been revised to clarify these issues and the stated position.
5020	Disagree. The reason this issue has reached regulatory levels is NOT because of a compelling animal model but stunning human observations. While these are scant in breadth (leading to all the problems elucidated above), to rely instead on a animal observations of a highly non-specific effect (whose relation to human BO is very questionable) would be disingenuous. Stay with humans. Deal with uncertainty with precaution as you have.	NIOSH agrees with the comment and this is the approach taken in the document. The TERA approach does not develop a proposed REL consistent with two important criteria: (a) 1/1,000 risk, and (b) working lifetime exposure. It would provide central guidance only in the absence of human epidemiological analyses. See also responses to comments 5086 and RA-45.
5021	I have limited the focus of my peer review to assess the technical validity of the information in the proposed criteria document and not matters of style or usage. I have specifically commented in my review on errors of fact, unsubstantiated claims, evidence of careless experimental work, inclusion of too much information already in the literature, or statements that are inaccurate.	No response needed
5022	In my review, I have placed special emphasis on the following issues and provide my review comments immediately following each. Following my review comment on these requested emphasis areas are general comments on other issues.	No response needed
5023	Based upon my reading of the proposed standard and my current knowledge and understanding of the exposure science addressing diacetyl and 2,3-pentanedione, NIOSH's recommendations in the proposed standard provides a reasonable and considered reflection of the scientific literature and state of the art for: Establishing causal link between occupational exposure and risk of disease; Establishing recommended exposure limits for diacetyl and 2,3-pentanedione; and Recommending methods of exposure control and assessment.	No response needed
5024	I am not aware of any.	No response needed

5025	Based on my experience and professional opinion they are reasonable but lack guidance in some specific areas that I believe the Industrial Hygiene and Occupational Health communities would find not only useful but also necessary. The recommended standard discusses the rationale for establishing the 8-hr REL for diacetyl and 2,3-pentanedione, however no discussion or recommendation is provided on how the 8-hr REL should be adjusted for longer shifts. Since many workers now work "non-standard" shifts this is important guidance that NIOSH needs to provide so that industrial hygienists can correctly adjust allowable exposure levels.	NIOSH has established RELs for full work-shift exposures as a means of preventing chronic health effects. The proposed NIOSH RELs for diacetyl and 2,3-pentanedione are established as 8-hour time-weighted averages (TWAs) for up to 40 hours in a workweek. NIOSH recognizes that for work shifts that exceed 8 or even 10 hours, and especially in cases where work schedules exceed a 40-hour workweek, the allowable TWA exposure for the day should be proportionately reduced according to prudent industrial hygiene practices.
5025a	The recommended standard discusses the fact that most "flavoring" exposures are not to a single compound but mixtures of compounds. The recommended standard needs to provide discussion and recommendation for assessing exposure levels to mixtures. For example, a worker could be exposed to diacetyl and a second flavoring chemical that could affect similar target organs or have synergistic effects thus making the REL of 5 ppb based on only the diacetyl exposure inadequate.	Addressing mixtures is important only when more than one component is contributing to the observed adverse effect. The document argues that diacetyl is by far the leading contributor in general industrial settings although there could be instances where other components such as acetoin are also making a contribution. The document presents the reasons why NIOSH believes the acetoin contribution was small in the G population. For mixtures with several contributing but uncharacterized components that are typically present together, from a regulatory perspective, choosing one better described component as a surrogate for the hazard is reasonable and often the only option. This was the case, for example, in the hexavalent chromium standard which applies to many forms of chromium, some in mixtures, and the coke-oven standard addresses coal-tar pitch volatiles, a complex mixture. Asbestos occurs in many physical forms, known to have variable associated risk. NIOSH recognizes the epidemiological challenge of

		characterizing risk in mixtures and this is a current research priority. An analysis of exposure response that examines only one contributing component would underestimate risk to the extent that the other components were distributed independent of the agent being analyzed.
5025b	The recommended standard discusses levels of recommended respiratory protection but doesn't address half-facepiece types of respirators used in conjunction with gas-proof goggles. Formaldehyde, as diacetyl, is an eye irritant and the OSHA respirator selection guidance for formaldehyde states, "A half-mask respirator with cartridges specifically approved for protection against formaldehyde can be substituted for the full-facepiece respirator providing that effective gas-proof goggles are provided and used in combination with the half-mask respirator.	NIOSH policy is to recommend only full facepiece respirators when there is the potential for eye irritation. Half mask respirators with goggles are not being recommended because NIOSH is not aware of any standards for gas-tight goggles that would permit NIOSH to recommend such goggles as providing adequate eye protection. This policy is from the NIOSH Respirator Selection Logic [2004c] page 21.
5026	My professional opinion is that in general they do. However, I am concerned about transparency with the issue of humidity corrections made to a large number of vapor exposure samples collected and analyzed using NIOSH method 2557. These exposures make up a significant portion of the worker exposure data upon which the recommended standard relies for its risk assessment and REL recommendations. I believe the research identifying the formula for humidity correction is valid work but it is not clear to me how the humidity data was collected to make the correction; my current understanding is that temperature and relative humidity data were not collected in the actual, indoor plant environment where the sampling was done but rather, general ambient outdoor temperature and relative humidity was used. I was not able to determine this from my review. Obviously that is an important detail	The temperature and relative humidity measurements used to calculate corrected values for environmental measurements of diacetyl with NIOSH Method 2557 were taken in the same baskets that were used for general area samples. Individual personal and area measurements were corrected before further statistical use, such as calculation of means. We have now included this information. The reviewer asked for an example of applying the correction to an actual exposure measurement of diacetyl. The inclusion of the Cox-Ganser et al. [2011] publication in Appendix 2 of the draft criteria document demonstrates the application of the correction procedure, especially in Figures 7 and 8.

	to clarify because in my experience interior temperature and relative humidity values can be markedly different from outdoor ambient levels especially in the months when the indoor environment is heated. I also think that providing an example of how the correction was applied to an actual exposure measurement of diacetyl would be helpful and be in the best interest of transparency. Again it is not clear whether corrections were made to individual samples or means.	
5027	I don't have the expertise to respond to this question.	No response needed
5028	I am not aware of any.	No response needed
5029	Other comments:	See comment 5029 a–g below.
5029a	Page 24, Line 15: Change "there" to these.	The document was revised as suggested.
5029b	Page 24 & 25, Line 19 P2 to Line 24 P25: Reference to method flowrate and time should use consistent units not be in ml/min one place and L/min in another.	The document was revised as suggested.
5029c	Page 31, Line 8: Change "involve" to "employ" or to "use."	Changed to "include the use of"
5029d	Page 38, Line 22: States the concentration was corrected but Table 1 doesn't indicate that the White et.al. data were corrected. Which is correct?	A footnote has been added to Table 1 for clarification.
5029e	Page 281, Line 12: Insert "to" between "exposure diacetyl."	The document was revised as suggested.
5029f	Page 283, Line 22 & 25: One sentence uses "OEL" the other "REL." In this context "OEL" should be used not "REL."	The document was revised as suggested.
5029g	Page 289, Line 27: The sentence beginning in line 27 with "However, no..." doesn't make sense as written."	This sentence was deleted.
5030	I don't have the expertise to respond to this question.	No response needed
5031	I don't have the expertise to respond to this question.	No response needed
5032	I don't have the expertise to respond to this question.	No response needed

5033	I don't have the expertise to respond to this question.	No response needed
5034	I don't have the expertise to respond to this question.	No response needed
5035	I don't have the expertise to respond to this question.	No response needed
5036	I don't have the expertise to respond to this question.	No response needed
5037	Thank you for the opportunity to contribute to this important NIOSH effort. I have organized my comments according to the Draft chapters, focusing specifically on Chapters 3, 5, 6 and 7.	No response needed
5038	This chapter [Chap 3] is comprehensive. NIOSH deserves much credit for its efforts to characterize the respiratory effects of diacetyl in exposed workers. The following comments suggest specific ways in which the worker study data could be more clearly summarized and more effectively presented in the Draft:	No response needed. This tracking number should be eliminated or combined with 5039.
5039	The chapter presents findings from numerous worker cohorts that sometimes overlap across reports. The actual data for some of the most important cohorts were presented in multiple reports; in some cases the various reports present the same data, but in other cases the data are inconsistent across the studies. It would be useful if the Document included tables to help readers "keep track" of the various reports and cohorts, thereby keeping track of their interrelationships. Table 3.1 (p. 47) is not adequate for this purpose.	To clarify Table 3.1, we have added facility alphabetical designations to the table. This allows identification of worker cohorts across publications.
5040	Some of the plants and workers were described in multiple reports and publications. It would be useful if NIOSH included a table that identified the various reports that presented data for the same plant and/or workers. Following are two examples: Example 1: Clinical findings in plant A workers were described in the following reports, each of which was cited at various places in the Draft: MMWR [CDC 2002], Kreiss 2002, Parmet 2002, Schachter 2002, Akpinar-Elci 2004, Kanwal 2006, NIOSH 2006, and Kanwal 2011. Example 2: In the Lockey (2009) report, one of the four plants described (plant #1) was also one of the plants included in NIOSH 2006 and Kanwal 2006 (plant F). This fact should be made obvious in the Draft. Interested readers should not have to spend as much time as I did to "decode"	We added the requested information in Table 3.1, as described above.

	the reports and thereby determine that these were in fact the same plant.	
5041	Some of the most important plants, especially those studied by NIOSH, are identified in the text using ID codes that differ across chapters. In chapter 3, the six popcorn facilities described in NIOSH 2006 and Kanwal 2006 are identified as plants A-F; A is also sometimes referred to as the "Index Plant". But in chapters 2 and 5, those same plants are identified using different ID codes. For example, chapter 5 reviewed data from plants A, D, E and F, but identified those plants as G, K, N and L. Those inconsistencies probably reflect different systems used within NIOSH (e.g., industrial hygiene vs. occupational medicine); similar differences are seen in reports published by NIOSH in the OEM literature (e.g., JOEM, AJIM) compared to the IH literature (e.g., JOEH). Such inconsistencies make the Draft unnecessarily difficult to read and understand. One approach to resolve this problem would be the adoption of a single set of ID codes used consistently throughout the Document. Alternatively the Document could provide a table indicating the various alternative names and codes used to identify each of the various plants discussed in the Draft.	The affected chapters have been revised to bring consistency to the citations of facilities as well as the references regarding those facilities.

5042	In some cases, NIOSH authors have published seemingly inconsistent results for apparently identical surveys and samples. Following is a prominent example: The reported mean diacetyl air levels for mixers and the number of samples obtained during the 2000 evaluation at plant A differ across reports: Report Mean Range # of Samples MMWR '02 18 not reported not reported Kreiss '02 32.27 1.34-97.94 12 samples Akpinar-Elci '04 32.3 not reported not reported Kullman '05 37.8 2.3-98 10 samples Kanwal '06 37.8 1.3-97.9 12 samples NIOSH '06 37.8 2.26-97.9 10 samples The above discrepancies are not the result of post hoc correction of earlier Method 2557 samples; these papers were all published prior to the adoption of the correction algorithm. I have not determined whether similar inconsistencies can be found for other results and/or other NIOSH reports. Such disparities might not impact the overall conclusion that diacetyl exposures should be restricted, but they might impact the calculation of appropriate exposure limits. This Criteria Document provides an important opportunity for NIOSH to present a "corrected" compilation of its own data. NIOSH should ensure that reported data from NIOSH studies are correct and presented consistently. Where there are inconsistencies across reports, NIOSH should indicate the "most correct" data and explain the reasons for discrepancies across its reports.	The draft criteria document's Chapter 2, section 2.5, gives the summary data for the NIOSH investigations, corrected for the underestimation of NIOSH Method 2557. The data appearing in Table 1 on pages 36 and 37 serve as the final "corrected" compilation of data. The Facility G data corresponding to the NIOSH [2006] and Kreiss et al. [2002] references for the initial cross-sectional study have been published with corrected values in Kanwal et al. [2011]. The discrepancies among the different publications offered by the reviewer are rounding differences in most circumstances. Where the number of samples differed (12 versus 10), later publications reflect reclassification of samples by work area, which changed the lower bound of the range from 1.3 to 2.3 ppm diacetyl, increasing the mean diacetyl concentration in mixing in Kreiss et al. [2002] from 32.27 to 37.8 ppm. The Kanwal et al. [2006] paper was based on previously published data (Kreiss et al. [2002] and Kullman et al. [2005]) and may have used these two sources, propagating an inconsistency. As the reviewer notes, these minor inconsistencies do not affect the overall conclusion that diacetyl exposures should be restricted, and the risk assessment was based on the latest "corrected" data from historical sampling. The summary job exposure matrix data in Appendix 3 of the draft criteria document are all based on corrected values and hence are updated from the NIOSH health hazard evaluation reports. This is now noted in Appendix 3 and in Chapter 5.
5043	The presentation of "adjusted" and "unadjusted" Method 2557 exposure data may confuse readers.	See response to comment 5045.
5044	The Draft uses two terms to refer to the transformed data: "adjusted" and "corrected". For clarity, and to avoid reader confusion, use of only one of those terms would be preferable.	See response to comment 5045.
5045	At various places, the Draft presents and discusses "adjusted" values, but cites references that reported only "unadjusted" values. For example (page 59, line 20): "Compared to mean diacetyl air concentrations measured at the index microwave popcorn plant, mean corrected diacetyl air concentrations at the other five microwave popcorn plants were lower: 0.02 to 0.83 ppm in the	The term "adjusted" has been changed to "corrected" when referring to the change in analyte concentration in response to humidity and analysis time. This should eliminate confusion in this regard. Other "adjustments" are mentioned in the document, such as an adjustment for smoking related to pack-year history, but these situations are clear in their own context.

	packaging areas and 0.63 to 1.54 ppm in the mixing rooms/areas [Kanwal et al. 2006]." Kanwal '06 contained only unadjusted data; the "corrected diacetyl air concentrations" are not found in that report. I suggest that when the Draft discusses the "adjusted" Method 2557 results from published studies that contained only "unadjusted" results, both the "unadjusted" data and the subsequently "adjusted" values be presented. For example, the mean adjusted area levels obtained during the first survey of Plant A mixers could be presented as: "[unadj: 32.3 ppm; adj: 57.2 ppm]".	
5046	The Draft does not adequately describe and utilize the Lockey 2009 data.	We have revised the description of the Lockey et al. [2009] data to include additional detail. The reviewer's comments regarding utilization of the Lockey et al. [2009] data are addressed in succeeding tracking numbers (5047, 5048, and 5127).
5047	NIOSH and Lockey independently studied workers at the same plant (NIOSH plant F, Lockey plant #1). The data could be pooled to allow a second longitudinal study, which would be a significant addition to the Draft (discussed further below, see 5-3).	The reviewer has misidentified the plant studied by both NIOSH investigators and Dr. Lockey. Facility L was studied cross-sectionally by NIOSH in 2003 (labeled Plant F in Kanwal et al. [2006]) and subsequently studied by Dr. Lockey in 2005–2006. The Lockey et al. [2009] paper aggregates data from four microwave popcorn plants owned by one employer and does not present plant-specific data. The reviewer suggests pooling the Lockey data and the NIOSH data from Facility L, presumably to extend the follow-up of the current workers studied by NIOSH in 2003 through January 2006, or even longer should Dr. Lockey's additional follow-up data through 2012 become available. Unfortunately, both NIOSH and the University of Cincinnati are precluded from sharing identifiable data that would enable linking. NIOSH is precluded from sharing its identifiable data with the employer, who hired Dr. Lockey. Dr. Lockey has shared a subset of deidentified data with OSHA but without pre-2005 exposure data on three plants and with correction for NIOSH Method 2557 measurements. A combination of pulmonary function measurements is unique enough to allow reidentification of employees by either party, should linkage be performed by a third party.
5048	Lockey et al. have recently completed a follow-up study of that plant; those additional data and their new longitudinal data should be reviewed for inclusion in the final Criteria Document. Review of	Lockey et al. have not completed their longitudinal analyses of the four microwave popcorn plants, which they plan to submit for publication.

	these data should be considered, even if a manuscript has not been yet published, because the data are particular importance given the limitations of the NIOSH longitudinal study (discussed below, see 5-1b and 5-1c).	
5049	This chapter [Chap 5] is comprised of two principal parts: 1) exposure-response modeling using data from plants G, K and L; and 2) a quantitative risk assessment. Those two parts and their limitations are tightly linked. There are insufficient data to support the low-dose extrapolations of the exposure-response modeling, which makes those extrapolations statistically and biologically uncertain. Because the low-dose extrapolations "may not reflect biological realities" [Crump, 1995] and because the number of studied workers is relatively limited, neither the epidemiological data nor the exposure-response models justify the very small benchmark responses (e.g., BMR = 0.1%) used in the human risk assessments. As a result, the BMD risk assessments presented in this chapter deviate importantly from standard methods, are not scientifically defensible, and should be revised.	The benchmark response value (BMR) = 0.1% is based on NIOSH precedent for 1/1,000 risk benchmark and non-cancer outcomes. The observations include a substantial proportion of the population exposed within a factor of 10 of the proposed REL. See also responses to comments 5050 and 5051.
5050	The exposure-response modeling is elegant in design. The insights provided regarding the likely presence of a susceptible subpopulation is especially interesting. However, the empirical data available for modeling are relatively limited; therefore the model conclusions are necessarily limited. This is of particular importance for the low-dose extrapolations.	Sufficiency of available data is frequently an arguable issue. In this study, measurements < limit of detection (LOD) are among the study's most precisely known values; they are very close to 0.0. Almost any imputation procedure will produce virtually identical regression results. See also responses to comments 5049 and 5051.
5051	A substantial proportion of the exposure data were <LOD; those values were set equal to LOD/2. Following are the proportions of <LOD data for the three plants included in the modeling (p. 116): G: 105/262 = 40.1% <LOD (personal samples), 46/346 = 42.2% <LOD (area samples); K: 44/60 = 73.3% < LOD; L: 4/125 = 3.2% <LOD. Assigning a value to <LOD samples is conventional, but the large proportions of arbitrary, imputed values at plants G and K raise concerns about the modeling results. Low-dose extrapolations cannot be made with confidence when >40-70% of total values, and essentially all of the low-dose values, are unknown. (I assume that 100% of "low-level exposures" were <LOD, depending on whether any were obtained over unusually long sampling periods). In such a	We do agree that model interpretations are limited but this issue is discussed sufficiently in the document. One additional point to be made is that the observations include a substantial proportion (> 20%) of the population exposed within a factor of 10 of the proposed REL. See also responses to reviewers' comments 5086 and EA-3.

	case, the dataset provides no information about the shape of the low dose dose-response curve and estimations of model variance would be biased downward by the arbitrary assumption that all the LOD samples had essentially identical values. This important limitation to the modeling exercise is not adequately discussed in the Draft.	
5052	The spirometry data utilized for "longitudinal analyses of ppFEV1 at Company G" were much more limited than might be inferred from the Draft. Chapter 5 states that 361 plant G (aka Plant A) workers were included in these analyses (Table 5.7): Group 1 workers hired prior to the initial survey and the implementation of exposure controls plus Group 2 workers hired after that survey and the implementation of exposure controls. But Chapter 3 and the original NIOSH studies indicate that the number of workers with multiple pulmonary function tests and sufficient follow-up duration to be appropriate for analysis was substantially smaller. The following is from NIOSH '06: "Unfortunately, because of high turnover of the plant workforce, only 86 (38%) of 227 workers hired after November 4, 2000 had more than one spirometry test by NIOSH; only 41 (18%) of 227 had more than two tests. These numbers were insufficient to provide stable or representative information about respiratory disease risk among newly hired workers." (p. 12) (emphasis added). Moreover, mean duration of employment in the Group 2 workers was only 6 months; the number of workers with two or more PFTs and one year of exposure is not described, but seems likely to be smaller. The above quoted concern about data insufficiency is consistent with a more general view expressed by NIOSH scientists and others that longitudinal studies using spirometry are "problematic" when fewer than three sets of PFTs are available for each individual (e.g., Wang & Petsonk 2004; Pellegrino et al. 2005; Berry: Bull Physio-path Respir 10:643, 1974). It seems a major concern that the Draft's dose-response modeling, risk assessment, and exposure limit recommendations rely primarily on data previously determined by NIOSH to be "insufficient to provide stable or representative information". The Draft should more clearly describe the limitations of the available data, indicating numbers of	The reviewer accurately states that the field investigators thought that the number of newly hired workers and their length of spirometric follow-up did not allow us to conclude that the lowered diacetyl exposures were sufficient to conclude that these workers had no respiratory risk. Indeed, the newly hired workers in high risk categories, such as mixers, did have excessive FEV_1 decline as shown in Table 4 of Kanwal et al. [2011]. This prompted our recommendation that workers continue in respiratory protection. With regard to the references the reviewer gave regarding the need for three or more sets of spirometry for longitudinal studies: This conclusion is dependent on data quality, test interval, and disease under consideration. Because of the variability of spirometric measurements, test intervals of many years yield more stable slopes of decline than numerous short-interval tests. However, adjusting for within-person variability and using population-based normative data for declines allow the use of short-interval tests (Chaisson et al. [2010]; Wang and Petsonk [2004]). Indeed, the Wang and Petsonk paper reports normal FEV_1 declines for two tests at either a 6-month or a 12-month interval. Thus the reviewer has overstated a concern about our use of the limited data for public health recommendations to Facility G. A major difference between the field investigators' approach in 2000–2003 and the later risk assessors' approach is in the interpretation of abnormal spirometry in workers hired subsequent to the initiation of interventions, which began in 2001. The field investigators were hesitant to interpret abnormal pulmonary functions at short tenure as work-related unless excessive decline in serial pulmonary functions had occurred. There were no preplacement (pre-exposure) pulmonary function measurements with which to assess whether abnormalities had evolved during employment. The low

	workers grouped by numbers of PFT exams and further sub-grouped by duration of employment. If NIOSH persists in relying on these workers as the basis for its dose-response models and risk assessment, then that should be reconciled with the 2006 characterization of those data as "insufficient". These critical issues have not been adequately discussed in the Draft.	proportions with serial measurements in the workers hired in 2001–2003 limited our early conclusions. In contrast, the risk assessors interpreted abnormal spirometry on first test for the 2001–2003 workers as work-related, without reference to serial declines. This interpretation may be reasonable because the latency for severe cases of pulmonary impairment was frequently a few months, a fact that we did not know early in our epidemiologic work. We have not changed the draft criteria document in response to this comment.
5053	a) In light of the limitations of exposure data and small numbers of longitudinal spirometry data, there does not seem sufficient "actual" (i.e., not imputed) data to allow for valid model building. If only 18% of post-2000 plant G (aka plant A) workers had ≥3 PFTs and if 40-42% of measured exposures were <LOD, then one might expect that only 10-11% of workers (i.e., 18% x 59% = 10.6%) had empirical data appropriate for valid modeling. (The actual number depends on whether there is an association between level of exposure and number of PFTs; the Document provides no information about this). b) One would expect that the low-dose extrapolations were especially dominated by data deficiencies; thus the low-dose exposure modeling probably relied on a vast preponderance of imputed, rather than empirical data. Accordingly, it seems likely that the low-dose extrapolations "may not reflect biological realities" (Crump 1995; see below). The limitations described above are not adequately discussed in the Draft, which should clearly detail the data limitations and the needs for caution in interpreting the low-dose extrapolations and related analytical results.	(a) This and other similar comments lead us to conclude that we did not sufficiently explain in the reviewed draft of the document that exposures < LOD are quite meaningful. In a regression analysis, these values are *extremely* meaningful; they represent low or zero-exposed workers whose experience is being compared to that of exposed workers, which is exactly the desired contrast. (b) The primary problem with exposures < LOD concerns the humidity correction, which would tend to introduce a great deal of variability and exposure misclassification. This in general would obscure the exposure response, not artificially heighten it. The revised draft of this criteria document contains considerably more description of both of these issues and reports a further analysis in which the humidity correction was applied to all samples, not just those above the LOD; the result was a very small change in the estimates of diacetyl effects.
5054	The Results tables indicate that various regression models had substantial statistical significance, probably reflecting a strong dose-response relationship in the range of objectively measured exposures. On the other hand, the R^2 values are relatively small, suggesting that despite the strong dose-response, the various models lack predictability. Moreover, the R^2 values are reported for	Most of the variability implied by the small R^2 arises from the inherent variability of lung function determinations and is not a reflection of uncertainty in the exposure response. The *P* values for the exposure metric estimates indicate quite well-defined effects. A t-statistic upon which the *P* value is based (and confidence limits

	the over-all models, not for the specific associations between lung function and diacetyl exposure levels. Thus, results of the regression modeling provide less precise information about the relationship between diacetyl exposure and altered lung function than might be otherwise inferred. This is of particular importance for the low-dose extrapolations, which are based on very few actual data. These limitations are not adequately discussed in the Draft.	could be based) was considered a more appropriate measure than R^2 for the exposure effect itself.
5055	a) Given the above noted limitations of the data from plant G [aka A], it is surprising that NIOSH did not perform a second longitudinal study by combining its data from plant L [aka F] and the data subsequently obtained at that plant by Lockey and colleagues [Lockey 2009; White 2010; White: J Occup Environ Hyg 8:D25, 2011]. Such a longitudinal analysis would provide perspective on and might validate the NIOSH longitudinal analysis of plant G. b) Table VII of White '10, which presents side-by-side comparisons of unadjusted sampling results from the Lockey '09 and NIOSH 2006 surveys, indicates similar exposure levels across those two studies. Thus combining the two studies would have been relatively straightforward. Moreover Lockey et al. have completed an extended follow-up and longitudinal study of the workers at that plant. NIOSH should either perform a longitudinal analysis combining its data and the Lockey data from plant L, or it should explain why such an analysis could not be performed.	(a) For production workers (the majority of workers with substantial exposure), average diacetyl exposures differed by a factor of almost 20 between Plants G (higher) and L (lower) as shown in Table 5.3. The criteria document discusses this and other reasons why Plant L was not the basis for the REL. (b) There are severe limitations in the Lockey data, including aggregated work histories and absent historical exposure estimates, which preclude combining the data. NIOSH will examine any additional Lockey data that may become available.
5056	The Draft notes that there was a "divergence in optimum exposure metrics" when the results of a cross-sectional analysis of pulmonary function at plant G [aka A] were compared to the corresponding analytical results of plants K and L [aka D and F] combined (p. 123). Besides the other substantial benefits of a second longitudinal analysis, combing the NIOSH and Lockey data from plant L would probably result in a database larger than that from plant G, which could provide insight into the selection of "optimum exposure metrics". The choice of "optimum exposure metrics" is critical to the dose-response modeling, the risk assessment, and the setting of exposure limits. Failure to clarify and explain that "divergence" may be a significant flaw in the modeling and a critical flaw in the risk analysis and standard setting.	The text explains that the exposure assessment at Plants K and L was very likely not representative of earlier diacetyl airborne concentrations, and that there were reasons to believe that exposure reductions had been achieved prior to the NIOSH exposure surveys. It provides some plausible explanations for the consequences of that inevitable exposure miss-classification on the observed exposure-response estimates using various metrics. This heterogeneity of the exposure-response estimates across the three plants (G, K, L) was an additional reason why a combined database was not used in the risk assessment (the other primary reason being the inadequacy of the retrospective exposure assessments themselves at Plants K and L).

5057	The exposure-response modeling did not consider the possible effects of short-term peak exposures. If short-term exposures exceeded upper airway scrubbing capacity, they would have had disproportionately large impact on the lower airways. Chapter 5 states that "peak exposures were not directly available … although selected jobs were analyzed using a real-time method" (p. 120). However, substantial peak exposures were documented in various NIOSH publications. For example: "Plant D [aka K]… had the lowest mixing area mean diacetyl air concentration. However …real-time monitoring in a mixer's breathing zone at plant D revealed peak diacetyl air concentration over 80 ppm over several minutes …" (Kanwal '06; p. 151) High peaks were mainly noted in mixing areas, but high transients were also reported for the QC job category. For example: NIOSH '06 (Table 3) indicates that diacetyl TWA air concentrations for QC workers at plant A ranged from <LOD to 0.02 ppm, but Figure 4 demonstrates real-time peaks of 7-14 ppm during that survey. Thus Group II QC workers at plant A may have had repeated peak exposures 350-700 times greater than their highest (unadjusted) TWA exposure levels. For such reasons, Kanwal '06 concluded: "Peak exposures may be hazardous even when ventilation maintains low average exposures". However, the exposure-response modeling ignored the possibility that repeated peak exposures caused significant effects despite low TWA exposure levels; the modeling and risk assessment considered only the TWA exposures. Because the effects of such repeated peak exposures in otherwise low-TWA workers were not considered, the resulting low-dose extrapolations are necessarily uncertain. These limitations are not adequately discussed in the Draft.	We have now performed analyses excluding all workers who were ever mixers, and the effect estimates are slightly *stronger.* This is strong evidence against the special vulnerability of mixers to peak values playing a dominant role in diacetyl-associated morbidity. There is no available study population that has documented historical time-integrated peak-exposure data for the entire population under study. The data do not support this speculation.
5058	Chapter 5 describes a risk analysis using "the widely used benchmark dose procedure" (p. 129), self-referring to "the conventional benchmark dose procedure" (p.129), and further stating that it had modified "the traditional BMD procedure" (p. 131) to include a susceptible sub-population. However, the risk assessment presented in the Document is not "traditional", "conventional" or "widely used" benchmark dose (BMD) analysis; it deviates from standard BMD methods and guidelines and it is also inconsistent with the	The self-referral was not to the statement "the widely used benchmark dose procedure" as implied by the comment but rather to examples of the application with continuous exposures. The document does not assert that the specific approach taken is widely used.

	methods recommended in BMD-related references cited in the Document. (The quotations in the following comments are from references cited in the Draft.)	
5059	The goal of BMD risk analyses is to provide an analytical method that avoids uncertainties associated with model-based low-dose extrapolations. Accordingly, it is standard teaching that the Benchmark Response (BMR) should be set at the lower range of the empirical data on which the risk assessment is based: Crump 1995: "The BMR is typically set at the lower end of the range of responses that can be detected experimentally, in order to avoid uncertainties associated with low-dose extrapolation using models that may not reflect biological realities." (p. 79) Clewell, 2003: "the most important criterion for selecting a BMR is that it should result in a BMD that is near the lower end of the range of the data providing information on the dose response. In particular, estimation of the BMD should not involve low-dose extrapolation, such as that traditionally performed in cancer risk assessments." (p. 1041) US EPA, 2000: "The major aim of benchmark dose modeling is to model the dose-response data for an adverse effect in the observable range (i.e., across the range of doses for which toxicity studies have reasonable power to detect effects)…" (p. 18) By contrast, the risk assessment in the draft Document used a BMR of 1/1000 (i.e., 0.1%), more than an order of magnitude below the experimentally observed range. Thus, the Draft BMD risk assessment is based on "low-dose extrapolation using models that may not reflect biological realities".	These comments all pertain to experimental animal studies in which there is a low dose group that represents the lower limit of observed exposure. The procedure is directed toward identifying a "point of departure" from which, typically, a linear extrapolation to the origin follows. In the diacetyl analysis, as noted by reviewers, there is sufficient low dose observation. The statistical models all implicitly extrapolate to the origin and we used a linear model for the REL, thus achieving the identical result that would follow from specifying a point of departure and linear extrapolation. Previously promulgated protective regulations specify a level that is orders of magnitude below levels where effects were being detected. Almost no epidemiological studies have ever detected excess adverse effects at exposure levels corresponding to the 1/1,000 lifetime risk level due to insufficient statistical power and too much measurement error and uncontrolled confounding.

5060	Standard (i.e., "conventional" or "traditional") BMD risk assessments use a BMR in the range to 1-10%: Crump, 1995: "Dose corresponding to 0.10 benchmark risk"... (p. 84) Clewell, 2003: "In the models for both quantal and continuous data the BMD was defined as the dose corresponding to 10% increase in the probability of an adverse response (i.e., as the solution to the equation P(BMD)−P(0) = 0.1) ..." (p. 1038) Park, 2006: "Similar to the impairment threshold issue, deciding what constitutes an acceptable increase in impairment prevalence, the 'benchmark response' (BMR), is another important question. An increment of 10% is often used in risk assessments…" (p. 374) The Park study ultimately used a range of BMR values: "1, 2, 5, or 10% for each outcome". (p. 377) US EPA, 2000: "A 10% response level is conventionally used (at least for dichotomous endpoints) to define effective doses … This response level is used for such comparisons because it is at the low end of the observable range for many common study designs, although for some designs the limit of detection is above the 10% level and for others it is below." (p. 19) US EPA Benchmark Dose Software (BMDS) On-line Tutorial, 2011: For Quantal Data: "An excess risk of 10% is the default BMR … If a study has greater than usual sensitivity, then a lower BMR can be used … For example, reproductive and developmental studies having nested study designs often have greater sensitivity, and for such studies a BMR of 5% has typically been used … epidemiology studies often have greater sensitivities and a BMR of 1% has typically been used for quantal human data." For Continuous Data: "If individual data are available and a decision can be made about which individual levels can be reasonably considered adverse … then the data can be "dichotomized" based on that cutoff value, and the BMR can be set as above for quantal data." Because the Draft BMD risk assessment used continuous data that were dichotomized, a "conventional" BMD risk assessment would have selected a BMR in the 1-10% range. The selection of a 1/1000 (i.e., 0.1%) BMR, rather than a conventional BMR of 1-10%, was a departure from standard BMD methodology. Such a BMR might be justified if it was "near the lower	This important study was exploring BMD applications for neurobehavioral effects. It described impairment arising in welders and other workers with various levels of manganese exposure over the course of a 2-year bridge construction project and did not propose a recommended level. Unlike the present application where there is a consensus definition of impairment (the lower limit of normal), impairment for neurobehavioral outcomes is less well defined. Impairment was defined as falling below various percentiles in a normal population, e.g., 1st, 5th, or 10th percentile. A subsequent published analysis of the same population with a more complete exposure history presented benchmark exposures corresponding to excess risks of 1/1,000, 1/100, 5/100 and 10/100 after a 2-yr exposure [Park et al. 2009]. With an acceptable excess risk stipulated a BMD can be calculated.

	end of the range of the data providing information on the dose response", but that was not the case for diacetyl.	
5061	Similarly, chapter 6 of the Draft Document describes concerns that a BMD analysis should not select a BMR below the range of observed data. The following statement describes this for 2,3-pentanedione: "BMD modeling was conducted for a response rate of 50% (BMD50) … rather than the conventional 10%, on the grounds that a BMD estimated from such limited data is less model dependent for points near the response rate of the observed data than at response rates requiring a greater degree of model-based extrapolation." (p. 204) That comment from the Draft reflects a more general concern: BMD analyses should employ BMRs "near the response rate of the observed data" in order to avoid uncertain "model-based extrapolation". The BMD methods presented in Chapter 5 ignored the BMD-related concerns described in Chapter 6. Such methodological inconsistency across chapters should be corrected or justified in the Draft. The BMD risk assessment presented in Chapter 5 deviated importantly from standard methods and guidelines, relied on low-dose extrapolations using models that "may not reflect biological realities", and ignored the guidelines and methods described in the references cited in that chapter.	The BMD modeling in the draft document for comparative potency analysis of 2,3-pentanedione, in comparison to diacetyl, has now been updated and replaced by a categorical regression analysis. This analysis is described in the revised Chapter 6, and is based on a conventional BMR of 10%. However, this should not be taken to imply that the lower BMR used in the epidemiological analysis is in any way improper. The epidemiologically-based risk assessment presented in Chapter 5 is based on far more data, including data at lower exposure levels, than the comparative potency analysis presented for 2,3-pentanedione. Furthermore, the two analyses have different objectives; the epidemiologically-based risk assessment for diacetyl attempts to estimate a safe level of exposure for workers, while the comparative potency analysis of 2,3-pentanedione attempts to estimate the toxicity of this substance relative to that of diacetyl. The difference in methodology of the two analyses is justified by both the characteristics of the data available for the analysis and the differing objectives of the two analyses.

5062	The available human data do not justify a BMR smaller than 1%. To the contrary, the actual empirical data suggest that a still larger BMR would have been more appropriate: Table 10 of NIOSH '06 indicates that the smallest observed "FEV1 declines of 300 ml and/or 10% from baseline in workers who had three or more spirometry tests" was 7%. Table 9 of NIOSH '06 provides cross-sectional data for "microwave popcorn packaging-area workers in Cohort-2 who participated in more than one survey", including data for "first survey" and "last survey". At the last survey, "obstruction on spirometry" was noted in 6.7%. Moreover, "there were no statistically significant changes in the prevalences of symptoms, spirometry abnormalities, or in mean percent predicted FEV1 from first to last surveys in Cohort-2 packaging workers." Based on such data, it could be argued that a BMR should be set at 5% for this risk assessment, with the resulting BMD used as a point of departure for extrapolation. In light of the published BMD guidelines cited in the Draft and discussed above, and in consideration of the limitations of the empirical data, NIOSH should reconsider its choice of BMR, provide justification for its choice, and perform a more standard and traditional risk assessment.	As mentioned in other responses, the BMR = 0.1% is based on NIOSH precedent for 1/1,000 risk benchmark and non-cancer outcomes. A BMR of 5% or 50/1,000 was determined not to be acceptable.
5063	It is possible that a non-standard BMD analysis was employed because the results of more standard analyses yielded exposure limits that NIOSH deemed to be insufficiently protective. Several of the references cited in the Draft note that the results of traditional risk assessments might lead to conclusions (e.g., exposure limits) inconsistent with the views of policy makers. Exposure limits derived by standard risk assessment methods might "seem" too high or be judged to pose "unacceptable" risks. Such value-based judgments should be distinguished from the risk assessments themselves. For example: Clewell, 2003: "In particular, estimation of the BMD should not involve low-dose extrapolation, such as that traditionally performed in cancer risk assessments ... This consideration is different, however, from the question of whether a particular benchmark risk for a specific effect is acceptable in a given population." (p.1041) Park 2006: "A benchmark response of 10% corresponds to a 100/1000 excess risk, which often could have	The risk assessment procedure was determined before policy implications of the unknown results were considered. A risk of 1/1,000 is a previously used ratio that has been used by both the scientific and regulatory communities, and is consistent with past NIOSH practice for epidemiologically-based quantitative risk assessments. "Standard risk assessments" that lead to a BMD implicitly embody value-based constraints; there is no objective definition of "acceptable." A correctly-performed risk assessment displays a range of risk over a range of exposure levels. By itself, there is no prescription of acceptable level.

	accrued in much less than a lifetime of exposure. For irreversible effects, this choice of BMR would be unacceptable in most public health settings."(p. 374) The challenge is to distinguish between conclusions derived from empirical data (i.e., the results of standard risk assessments using observable exposures and observed effects) and policy decisions based on hypothetical data (i.e., risks estimated by extrapolations far outside the observable data range). To the extent that correctly performed risk assessments yield results inconsistent with policy considerations (i.e., risk assessments leading to exposure levels that "seem" inadequately protective or otherwise "unacceptable"), both the results of the correctly-performed risk assessments and the related policy considerations should be explicitly and independently presented. The risk assessment and its use for policy development should be presented as completely and transparently as possible.	
5064	This chapter [Chap 6] is comprised of four principal parts: 1) dose-response modeling of the Morgan mouse data, with focus on a comparison of the EPA default approach vs. a computational fluid dynamics (CFD) approach to dose metric calculations; 2) calculation of human equivalent doses; 3) BMD risk assessment for diacetyl; 4) BMD risk assessment for 2,3-pentanedione. The chapter relies extensively on the Allen "contractor reports", which are presented in Appendix 5. The Draft asserts that it contains "a summary of the risk assessment extracted from these reports", but it actually contains numerous pages taken verbatim from the Contractor Reports.	No response needed
5065	**Dose-Response Modeling** The dose-response modeling is thoughtful. I have no comments on the specific choice of CFD vs. default approach. I am concerned, however, by the assumption that: "no matter what region … is affected in mice (nasal or TB) …the dose-response relationship in humans is proportional regardless of the affected region in humans … no site concordance is assumed" (p. 181). This assumption is convenient for modeling, but no evidence is presented to support that assumption in either the Draft or the Allen Reports. The implications of this assumption are not adequately discussed in the text.	The toxicologically-based risk assessment for diacetyl has been updated on the basis of a new National Toxicology Program (NTP) study that was not available at the time that the draft document was written and no longer relies on the Allen report. The updated analysis does continue to assume that toxicity observed in the upper respiratory tract of mice is relevant to the estimation of bronchial and pulmonary toxicity in humans. Computational fluid dynamic modeling suggests that a larger fraction of an inhaled dose of diacetyl will reach the deep lung in humans than in rodents; therefore, it is not surprising that the most sensitive respiratory target region in mice may differ from that observed in humans. The

		toxicologically-based risk assessment is predicated on the assumption that the tissue dose of diacetyl that causes respiratory toxicity in mice may also cause respiratory toxicity in humans, even if that toxicity is observed in a different region of the respiratory tract.
5066	**Calculation of Human Equivalent Doses** The Draft presents two alternative approaches to calculating the HED, "tissue concentration" and "regional penetration". The authors of the chapter could not determine which was preferable, so both were presented. That approach is understandable, but unsatisfying. As presented in Table 6.4, the two HED approaches resulted in estimated BMD and BMDL values that varied by 10- to >100-fold as BMRs ranged from 10% to 0.1%. The largest differences were seen at smallest BMR values (i.e., BMD = 0.001), underscoring the model uncertainty due to low-dose extrapolations.	This comment pertains to the toxicologically-based risk assessment presented in the draft document. That analysis has been updated and replaced by a categorical analysis, using a new NTP study that was not available at the time the draft document was written.
5067	**BMD Risk Assessment for Diacetyl** The BMD analysis in Chapter 6 is closer to a "conventional" or "traditional" analysis than is the risk assessment in Chapter 5. For example, it uses the US EPA Benchmark Dose Software and follows the basic methodologies set forth by EPA. Like the diacetyl BMD analysis in Chapter 5, however, the Chapter 6 BMD analysis deviates from methodological standard by its use of a BMR = 0.1% (i.e., 1/1000), a value far below the range of observable effects.	The BMD analysis presented in Chapter 6 of the draft criteria document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000.
5068	1. As discussed above, a standard BMD analysis should set BMR at the lower range of the empirical data on which the risk assessment is based. In this case, the BMR of 0.001 is far below that "lower end of the range of the data providing information on the dose response" in the Morgan study. More specifically, as discussed in Chapter 6 and the Allen report: "The limited number of observations (5 per dose group) [and] the relatively small number of doses examined … contribute to the general uncertainty." (Appendix 5, p. 22). From this data set, the smallest observable dose-related effect would have been 0.2 (i.e. 1/5), not 0.001 (i.e., 1/1000). The choice of BMR = 0.1 (i.e., BMR = 10%) might be acceptable, but BMR = 0.001 (i.e., BMR = 0.1%) is not because it is more than an order of magnitude below the range of experimentally observable outcomes.	The BMD analysis presented in Chapter 6 of the draft criteria document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000.

5069	2. The NIOSH decision to analyze the Morgan mouse data using BMR = 0.001 apparently reflects the method used in the Allen "contractor reports". However, that method deviates from standard BMD methodology. I was interested to know whether any of Allen's published peer-reviewed reports had used and/or advocated setting such a BMR for BMD analysis. To determine that, I used the NML databases to identify Allen's published work; I then reviewed those that contained or described BMD analyses. I found no report in which he had used, advocated, or justified BMD methods using a BMR far below the range of responses detected experimentally. In fact, all of his published BMD analyses used "traditional" BMD methods, setting BMR in the range of 1-10% (i.e., 1/10 to 1/100); none used or advocated BMR = 0.001. Thus, the BMD methods used in the Allen "contractor reports", in particular the setting of BMR = 0.001, deviate from standard BMD methods and are inconsistent with the methods described in Allen's peer-reviewed publications. Accordingly, the Draft should not rely upon the Allen "contractor reports" as justification for setting BMR = 0.001 in its BMD analyses.	The BMD analysis presented in Chapter 6 of the draft document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000.
5070	3. Both Chapter 6 and the Allen "contractor reports" cite a paper by Wheeler and Bailey as support of their methods. However, that paper does not support setting BMR = 0.001 for the Morgan dataset. To the contrary, Wheeler and Bailey recommend the use of a traditional BMR in the range of observable data: Wheeler & Bailey 2007: "… the benchmark response (BMR) is commonly set at values of 1%, 5%, and 10%, where these values can be thought of as risks consistent with responses that are typically observable within the range of the data." (Wheeler p. 660) "… we are interested in estimating the dose associated with some specified excess risk for BMRs of 1% and 10%." (Wheeler, p. 664) Thus the BMD analyses in Chapter 6 and in the Allen contractor reports ignore recommendations in this reference and in others that were cited in the Draft (discussed above, see 5-6a and 5-6b).	The BMD analysis presented in Chapter 6 of the draft document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000.

5071	4. In the next section of this chapter, the 2,3-pentanedione risk assessment, the authors caution against setting a BMR far below the lower end of the range of responses that can be detected experimentally: "BMD modeling was conducted for a response rate of 50% (BMD50) … rather than the conventional 10%, on the grounds that a BMD estimated from such limited data is less model dependent for points near the response rate of the observed data than at response rates requiring a greater degree of model-based extrapolation." (p. 204) But that reasonable view of caution was ignored in their diacetyl risk assessment. Besides representing non-standard risk assessment methods, it reflects inconsistency across this chapter which should be corrected.	The BMD analysis presented in Chapter 6 of the draft document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000.
5072	5. The Discussion section cites the 2008 TERA (IDFA) risk assessment, but neglects the subsequent peer-reviewed publication: Maier et al: Evaluation of concentration-response options for diacetyl in support of occupational risk assessment. Reg Tox Pharmacol 58:285-296, 2010. BMD Risk Assessment for 2,3-Pentanedione The risk assessment for 2,3-Pentanedione is based on too little data to provide more than a superficial sense of its toxic potential. Ultimately, its assessment here is based on analogy, mainly structure-activity relationship with diacetyl and head-to-head comparisons of 2,3-pentanedione vs. diacetyl which are limited by small numbers of animals, differences in experimental protocols, and the use of differing species and strains. The analyses in this section are reasonable for hypothesis generating and the initial characterization of potential toxicity, but they do not seem appropriate or sufficient to support the setting of objective exposure standards.	The Maier et al. [2010] risk assessment for 2,3-pentanedione was included in the update of Chapter 6. NIOSH agrees that the comparative potency analysis of 2,3-pentanedione is based on a small study with few animals. NIOSH acknowledges that the characterization of the potential toxicity of 2,3-pentanedione is preliminary in nature; however, NIOSH considers that the data are sufficient to develop an initial recommendation for exposures to this compound. NIOSH notes that the analysis is based on the 95% lower-bound estimate (BMDL) of the benchmark dose (BMD) for 2,3-pentanedione. It is a characteristic of the benchmark dose method that the statistical variability of the data due to the small number of animals is reflected in the BMDL. NIOSH considers this adjustment for the variability of the data, by relying on the BMDL rather than the central estimate of the BMD, to be appropriate in setting a recommended exposure limit on the basis of preliminary data. The BMDL-based recommended exposure limit is intended to protect workers who are being exposed to 2,3-pentanedione while further toxicological testing is being conducted. NIOSH recommendations on exposure to 2,3-pentanedione will be updated when additional toxicological or epidemiological dose-response information becomes available.
5073	This chapter [Chap 7] provides a concise summary of Draft findings and recommendations. Unfortunately, it does not present a "transparent and sound basis" for standards.	The basis for the standards goes back to Chapters 5 and 6. We have revised Chapter 7 to address the concern for further transparency.

5074	Because most of the studies on diacetyl have focused on BO, and because most of the human and animal data discussed in the Draft focus on BO, it is surprising to realize that the proposed REL is intended primarily to prevent decline in FEV1, not BO. NIOSH should be clear and consistent in explaining its focus. For example, Chapter 7 contains the following three statements: "NIOSH has concluded that worker exposure is associated with a reduction in lung function" (page 212, line 20); "The NIOSH objective … is to reduce the risk of decreased lung function and the severe irreversible lung disease constrictive bronchiolitis obliterans" (page 213, line 2); "In occupational exposure to diacetyl, the health effect of concern is bronchiolitis obliterans" (page 213, line 13). As reflected in the submitted public comments that I reviewed, numerous stakeholders do not understand the focus of this Draft. (Likewise, it seems not to be adequately understood by a number of the occupational physicians, with whom I have discussed this over the past months). NIOSH should more clearly explain that evaluations initiated by the finding of unexpectedly high rates of BO in diacetyl-exposed worker led to evidence of accelerated declines in FEV1 in numerous other workers at those workplaces. And, because accelerated declines in FEV1 are important health effects that should be prevented and because exposure limits likely to prevent declines in FEV1 are also expected to prevent BO, NIOSH is proposing standards to prevent obstructive airway effects. It should be clear that the health effect of concern is decreased lung function, more specifically accelerated decline in FEV1, not bronchiolitis obliterans. Exposure levels set solely to prevent BO would almost certainly be higher than those set to prevent accelerated decline in FEV1, as proposed and discussed in the Draft.	A statement has been added to the introduction of Chapter 7 to make the basis of both the diacetyl and 2,3-pentanedione standard more clear. Additionally, section 7.5.1 states the REL for diacetyl is established that workers "should have no more than a 1/1,000 excess risk of suffering from reduced lung function associated with diacetyl exposure."
5075	Because the Draft risk assessments used nonstandard methods (see above, 5-6), the actual recommended exposure limits should be recalculated.	See response to comment 5071 and 5072
5076	I suspect that the exposure levels derived by standard risk assessment methods, using both the human and animal data, would have resulted in exposure limits that, in the judgment of NIOSH, "seem" too high or pose "unacceptable" risks. Such judgments may	The risk assessment performed by NIOSH using health hazard evaluation data was performed with absolutely no prior expectation or target for the resulting recommended REL (on the part of those performing the risk assessment). Subsequent to

	in fact be reasonable and even "correct", but they are not based on sufficient objective data to be tested and verified. NIOSH should distinguish between conclusions derived from empirical data (i.e., the results of standard risk assessments) and those driven by policy considerations.	establishing the range of lifetime risks, the risk assessors did not observe any policy imperative other than concern over levels of measurability; if anything, there was concern that the methods used were producing estimated risks that were too high, implying RELs that were too low.
5077	I appreciate concerns that 2,3-pentanedione might cause lung function declines in workers similar to those caused by diacetyl, but there are apparently no human data and only very limited animal data. The Draft is clear about such limitations: "Published reports on the toxicity of 2,3-pentanedione suggest that in rats 2,3-pentanedione causes airway epithelial damage …" (page 214, line 27) "… the current toxicological data for 2,3-pentanedione are limited, some preliminary conclusions … " (page 215, line 9) Such "suggested" effects and "preliminary conclusions" are adequate to justify warnings and cautions, and they emphasize the need for further research, but they provide insufficient objective, scientific data to justify a formal REL.	In addition to 2,3-pentanedione's structural similarity to diacetyl, the available toxicological data suggest that like diacetyl, 2,3-pentanedione can cause airway epithelial damage [Hubbs et al. 2012; Morgan et al. 2012]. The studies for the basis of the 2,3-pentanedione REL are summarized in Chapter 4 of the revised criteria document. In addition, Chapter 6 of the revised criteria document provides a relative potency estimate of 2,3-pentanedione, compared to diacetyl, indicating that 2,3-pentanedione may have equal or greater toxic potential relative to diacetyl.
5078	As noted in my prior comments, a major challenge to the exposure-response modeling is the relatively limited amounts of data. The challenge begins with the large numbers of <LOD samples, especially at plants G and K: G: 105/262 = 40.1% <LOD (personal samples), 46/346 = 42.2% <LOD (area samples); K: 44/60 = 73.3% < LOD; L: 4/125 = 3.2% <LOD. There may be little meaningfulness to model results based on such limited data sets, especially with respect to the modeled low-dose extrapolations. Question #1 concerns the "utility" of combining those three data sets. Further combining those sets would increase the number of samples, but it would not provide much additional information about low-exposure dose-response because there would still be essentially no data describing the low-dose exposure range. A different concern is the apparent heterogeneity across plants and their corresponding data sets, as discussed in the Draft and detailed in the Additional Questions. If the plant G data set is inherently different from those of plants L and K, then combining those heterogeneous sets would be unlikely to provide much insight into the underlying biological processes, even if the larger number of observations resulted in apparently greater	See response to comment 5097.

	statistical significance. It is possible that the resulting heterogeneity would more than offset the benefits of pooling the data. It would seem more useful and logical to combine the NIOSH data from plant L (aka F) with the Lockey et al. data from their multiple assessments of that same plant. That would provide a larger dataset with relatively little heterogeneity.	
5079	This question concerns which of three spirometry-based criteria should be used to define pulmonary obstruction. Each of the three alternative case definitions has benefits and limitations. It is my understanding that the high quality and consistency of NIOSH spirometry data would allow all three of those definitions to be used; by contrast, "real world" worksite spirometry programs would be likely to yield less reliable results for FEV1/FVC. If the question posed is whether NIOSH should use FEV1/FVC < LLN as an additional case definition for modeling its own HHE survey data, then the answer is "yes". The advantage is that the use of the ratio alone might allow detection of a few early cases of obstructive disease that had not yet progressed to FEV1 < LLN. The potential disadvantage is that use of the ratio alone might yield false positives, e.g., in very physically fit individuals.	Reviewer C suggests using an abnormal FEV_1/FVC ratio as an alternative outcome to using abnormal FEV_1 or the combination of abnormal FEV_1 and abnormal FEV_1/FVC ratio (case definition 2 in Table 5.37). Dr. Park notes that an isolated FEV_1/FVC abnormality will miss persons with restriction, although it may be more sensitive for early obstruction. We did not change the draft criteria document in response to this comment, as the possible increased sensitivity for early obstruction has to be balanced against misclassification, as pointed out by the reviewer.
5080	The Draft states (p. 121): "in identifying cases, a date of onset for a condition resulting in impairment and possibly representing BO was estimated as the average of the dates …"; such cases were included in the exposure-response modeling. By contrast, asymptomatic workers with abnormal spirometry were not included in the analyses. This Additional Question asks whether the exposure-response analyses should also include those asymptomatic workers. This question raises an issue noted in public comments and in my prior comments: Is the Draft about BO or about spirometric changes? Inclusion of asymptomatic subjects with impaired PFTs might contribute to an understanding of spirometric changes, but it would not necessarily contribute to knowledge regarding BO, at least to the extent that BO is regarded as a manifest clinical disease.	The revised Chapter 3 now includes a discussion of the insensitivity of spirometric abnormalities in pathologic constrictive bronchiolitis derived from observations of U.S. soldiers in Iraq and Afghanistan with exercise limitation (exertional shortness of breath), Iranians after mustard gas exposure, and a clinical case series. (However, we have not undertaken risk assessment analyses based on onset of shortness of breath alone.) We have little information about pathologic constrictive bronchiolitis in asymptomatic workers, but the degree of excess obstruction in the sentinel Facility G would suggest that obstruction associated with occupational flavoring exposures can be asymptomatic. We have also now included in the revised Chapter 3 evidence that constrictive bronchiolitis can result in a range of spirometric abnormalities, including restriction. Hence,

	(The Draft does not present evidence that BO develops insidiously from relatively persistent, asymptomatic airway obstruction or that BO exists as an asymptomatic condition). Including the asymptomatic subjects would move the Draft further from the "index cases" and BO, and further towards non-specific obstructive lung disease. If this is done, the Draft should be revised to emphasize that spirometric change, not clinical BO, is the focus. An additional concern is that the current inclusion criteria provide a reasonable basis for assuming that the observed spirometric changes developed after the start of diacetyl exposure. Inclusion of asymptomatic individuals would raise concerns that spirometric changes existed in some workers prior to diacetyl exposure. If NIOSH decides to include asymptomatic plant G workers with abnormal spirometry in these analyses, it should consider including only cohort 2 workers because there would be less uncertainty about levels and duration of historical diacetyl exposure.	our reliance on spirometric abnormality as an outcome for quantitative risk assessment makes sense. Given that many of the workers with obstructive abnormality were asymptomatic, both in the sentinel Facility G and in California medical surveillance, our risk assessment based on spirometry abnormalities (without inclusion of asymptomatic workers with obstruction) likely suffers from underestimation of effect. One option for including the asymptomatic workers with obstruction in the risk assessment would be to assume that contributing exposure either continues until spirometric abnormality was diagnosed or for half of the exposed period, if the risk assessors prefer. We based the risk assessment on spirometric abnormality rather than on constrictive bronchiolitis because we have very poor clinical tests for constrictive bronchiolitis until impairment is severe, and field studies do not include clinical diagnosis. Biopsy is insensitive in the hands of many pathologists. Paired high resolution computerized tomography scans were insensitive in the case series compiled in California flavoring worker surveillance [Kim et al. 2010]. Spirometry has been shown to be insensitive in biopsy-confirmed case series [Ghanei et al. 2008; King et al. 2011]. The reviewer is correct that we do not have evidence that persons with mild spirometric abnormalities progress to bronchiolitis obliterans in these industries, although we do have case reports of persons who progressed from normal to substantial impairment within months of exposure [Akpinar-Elci et al. 2004; Israel et al. 2009; Kreiss et al. 2002; NIOSH 2008b]. We do not agree with the reviewer that we should only include asymptomatic Facility G workers in cohort 2 because of enhanced certainty about levels and duration of historical diacetyl exposure. The workers in cohort 2 were not tested preplacement with spirometry. Thus, their abnormalities may or may not be related to work exposures.

5081	This is an Interesting question. By definition, CumExp and Dur interact (Cum Exp = Dur x AvgExp), but they are not necessarily collinear. To the contrary, the data presented in the Draft suggest that they have opposite signs. The challenge is to explain that relationship. The Draft proposes varying susceptibility (and depletion of a susceptible sub-population) to explain both the negative coefficient for duration and the "relatively strong but implausible prediction" that average exposure determines obstruction. This explanation is plausible, but unproven. As presented in the Draft, this finding is based on post hoc analysis of a relatively small dataset representing essentially only one plant (plant G). (The L + K analysis included a very small number of affected workers identified using only the less "specific to obstruction" case definition). An alternative possibility involves the likely presence of scrubbing and detoxification mechanisms. To the extent that those processes reduce the "effective" tissue dose (as compared to the external ambient dose), they would also reduce the probability of injury and disease. They would be most protective for low-level exposures, as compared to high-level and peak exposures that exceeded intrinsic scrubbing and detoxification capacities. Thus, transient high-level exposures are likely to cause more disease than long-duration low-level exposures, even if they resulted in the same cumulative external exposures (as determined by IH methods). That might explain the negative coefficient for duration seen in the models. Such a possibility was discussed in an earlier NIOSH publication: "Peak exposures may be hazardous even when ventilation maintains low average exposures" [Kanwal '06]. If this is correct, then the correlation between Dur and CumExp would depend on the level and pattern of IH-determined workplace exposures. For any given CumExp, greater correlation would be seen for low level exposures with few peaks, and lesser correlation would be seen when there were frequent high-level peaks.	The reviewer has offered a plausible explanation for the unusual findings of the risk assessment that cumulative exposure and duration have opposite signs in the models. Indeed, the high risk jobs in all plants appear to be ones in which peak exposures or high exposures occur, as the reviewer points out by citing Kanwal [2006]. The early onset of disease during employment curtails the accumulation of cumulative exposure, as reflected in the Akpinar-Elci [2004] paper in which the sentinel former worker cases had a median of 1.5 years of employment. We also noted that low average exposure was associated with excessive risk in quality control workers in company G, which may be a reflection of intermittent peak exposure. We have revised Chapter 3 to include information on the possible significance of peak exposure as a risk factor on pp. 3-12 and 3-17. We have revised Chapter 9 to indicate that peak exposures, not addressed in environmental surveillance, are an additional justification for medical surveillance (p. 9-1).

5082	This question asks about alternative exposure metrics for the BMD risk analysis. As noted in my prior comments, the Draft BMD analysis deviates from standard methods. The refinements proposed in this question might lead to small changes in risk estimates, but those changes would be small compared to the magnitude of the distortions caused by the nonstandard risk assessment method. If the BMD analysis was properly performed, focusing on data in the observable range, the impact of such refinements (e.g., cumulative vs. TWA exposures) could be readily determined. It is not obvious whether either of those exposure metrics would lead to significantly different POD values. If not, then it would not matter which metric was used. On the other hand, if they led to significantly different PODs, and if there were no other basis for choosing between them, NIOSH could select the exposure metric that yielded the lower POD and, therefore, the more protective exposure limit.	See response to comment 5018.
5083	The presence of a sufficient number of susceptible individuals in a potentially exposed population would be an important consideration in setting exposure limits. It would also be important to understand the biological basis for such susceptibility and its population prevalence. As noted previously, there are relatively limited data from which to conclude that there is a high-risk subpopulation of diacetyl-exposed workers. That possibility deserves further evaluation, but the current database is insufficient to justify including such susceptibility as a key element in developing an exposure limit. The limitations are several-fold. The data presented in the Draft are almost entirely from cases identified by case definition 1, i.e., the case definition less “specific to obstruction”. In addition, the possible adverse effects of transient peak exposures have been ignored; if such exposures overwhelm the scrubbing and detoxification capacity of the upper airways they are likely to cause disproportionate impact on the lower airways. Such transient peaks could explain the occurrence of pulmonary impairment after relatively short exposure duration and relatively limited cumulative exposure. To be useful and credible, the model should consider peak exposures (which NIOSH has documented to occur), rather than only cumulative exposures. In addition, the findings should be confirmed	See response to comment 5019.

	using the more "specific" case definition, and the findings should be validated in at least one other worker population.	
5084	I favor the use of human over animal data, when the human data are sufficient. In the present case, I would most prefer a human-based risk assessment that considered both the longitudinal data from plant G and the data from a pooled longitudinal study of plant L plus the Lockey et al. data. One challenge to using the animal data is the inability to determine the best approach to calculating HED values. I also have concerns about the NIOSH assumption that there is "no site concordance to be assumed" when considering animal lesions and corresponding effects of the human respiratory tract; this assumption is analytically convenient, but not adequately justified. With respect to the NIOSH and TERA analyses using the Morgan mouse data, I suspect that the differences in estimation of regional penetration and HED determination could be resolved and harmonized. (It is interesting to note that for any given BMR, the NIOSH analysis leads to a higher exposure level than the TERA analysis). The more significant difference between them is the TERA use of BMR = 10%, a value 100-fold greater than the BMR used by NIOSH. As should be clear from my prior comments, I disagree strongly with the NIOSH choice of BMR. A simple compromise would be to use the TERA risk analysis, which yields a more protective exposure level than does the NIOSH use of those data, but to choose a smaller BMR (e.g., BMR ≈ 5%), which is justified given the numbers of mice in the study, and would yield a POD and exposure level that would be more worker protective. Ultimately, given the limited available human data and the uncertainties regarding the best approach to using the animal data, ensuring the use of standard risk	See response to comment 5020.

	assessment methods is more important than the choice between the human vs. animal data.	
5085	As one of the reviewers as to comment of the proposed criteria document for diacetyl and 2,3- pentanedione I have reviewed the document, attended the Washington briefing on these matters held by NIOSH in August 2011, and read all of the submitted written comments to the NIOSH Docket Office regarding this subject. There was a charge to reviewers and this has been followed in my assessment of the materials that I have read. What was striking, but perhaps not surprising, about the public comments was the diversity of opinion. Given the various interest groups on various sides of this matter it was not surprising to see some commentary denigrating the NIOSH activities to date, and feeling that no such criteria document should be forthcoming at the present time. There were some legitimate concerns, such as the finding that diacetyl is a naturally occurring material in food products but this can be dealt with in a manner that does not prevent an appropriate set of controls being put in place for this material and related compounds. There are also some issues that need to be reflected upon, namely, that significant additional research in some areas of this question should be undertaken. First and foremost there should be better laboratory methods for identification and measurement, and the issue of "related compounds" should be made more clear.	We agree and have revised the discussion of analytical issues in Chapter 2 accordingly.

5086	This reviewer also takes issue with the concept that 1in 1,000 getting all is "acceptable". This reviewer finds this unacceptable for a variety of reasons. First, there are some 1.5 million food manufacturing workers estimated at the current time to be active in the United States. While all may not handle diacetyl and related compounds there are many who do and 1in 1,000 means many people will potentially get sick. Even more telling, there are some 6,500 laboratory workers who are involved with flavor manufacturing or laboratory activities and even among this small group that 6 individuals getting sick would be considered by this reviewer intolerable. However, given that there is at present poor laboratory assessment for diacetyl and related compounds this criteria document could well be put in place as recommended, perhaps with some minor changes, and revised in the near future when better laboratory methodologies become available.	This is a policy issue that was determined during the development of this document. At least 14 peer reviewer and 7 public comments raised issues regarding the BMD procedure used in the risk analysis for this document. This set of critiques conflates two issues: (a) what is an appropriate maximum acceptable level of excess risk, and (b) what is a valid BMD procedure for human population data with continuous exposure and outcome metrics. Current policy is driven by two consequences of the OSH Act: (a) no material impairment in a working lifetime, and (b) in the matter of the benzene standard, the Supreme Court conclusion that one in 1,000 excess deaths attributable to work in a lifetime is too high. Standard risk assessments that lead to a maximum acceptable exposure implicitly embody value-based criteria; there is no objective definition of “acceptable risk.” A correctly performed risk assessment displays risks over a range of exposure levels and does not, by itself, prescribe an acceptable level. Defining respiratory impairment using the LLofN provides a basis for calculating risk. Applying the criterion that excess risk not exceed 1/1,000 implies not only that at most one additional individual will fall below the LLofN (become impaired) per 1,000 workers, but also that the increase in impairment will be limited for those already below the LLofN (approximately 100 out of 1,000), a group that is assumed to have the same susceptibility as others. An excess risk of 5% or 10% (50 or 100 per 1,000) would not be acceptable; substantial numbers of those already below the LLofN would become much more seriously compromised [Bellinger 2004]. Almost no epidemiological studies, even of smoking in entire countries, have observed statistically significant excess adverse effects at exposure levels corresponding to 1/1,000 excess lifetime risk; there is insufficient statistical power, too much measurement error and uncontrolled confounding to permit that. All of the criteria document reviewers’ comments citing the standard benchmark dose methodology pertain to experimental animal studies. When dealing with discrete outcomes (impaired or diseased: yes/no) typically a point of departure (POD) is established, in one of two ways: (a) there is a low dose group that represents the lower limit of exposure with

		observed adverse effect (the LOAEL), or, (b) the procedure fits a statistical model and then selects for the POD a low exposure level at which there is still a robust estimate of the probability of impairment – within the range of the observed data, and typically representing a 5% excess. From the POD, traditionally, either (a) uncertainty factors are applied with the intent of obtaining an equivalent estimate in humans accounting for interspecies and other variability, or (b) a linear extrapolation to the origin from the POD is made to identify a dose (the BMD) corresponding to the specified maximum acceptable excess risk (the BMR). For impairment as a discrete outcome, excess risk estimates can be derived and a BMD calculated directly. In this risk assessment, we chose to model loss of pulmonary function as a continuous variable (e.g., percent of predicted FEV_1) with multiple linear regression, rather than as a discrete outcome (impaired: yes/no) using multiple logistic regression. The latter method would have permitted direct estimates of excess risk and BMDs, but probably with less statistical power because less information is being used (36 out of 360 cases vs. 360 individual FEV_1 measures). However, following the POD tradition in the current risk assessment one could choose as a POD the exposure 0.2 ppm, which is well within the range of the observed data and an exposure very credibly associated with health effects. Applying in the traditional fashion an uncertainty factor of 10 for interindividual variability, and an uncertainty factor of 3 for comparing exposure duration in the study population to a 45-year duration, one obtains a "safe level" of $0.2 \times 0.1 \times 0.33 = 0.0066$ ppm, which is close to the proposed NIOSH REL. This procedure does not estimate an associated level of excess risk and was not the basis for the REL. In BMD applications for continuous outcomes, a POD is not necessarily needed or used. For example in an animal (mice) study of pentabromodiphenylether, a flame retardant, the effect measure (locomotion) was modeled and the BMD derived directly (a) as the exposure for which there was a 5% or 10% loss of function (not an estimate of risk), and (b) as exposures conferring 5% or 10% increased risk using two definitions of impairment [Sand et al. 2004]. In a study of child test scores related to mercury levels

		in cord blood or maternal hair, [Budtz-Jørgensen et al. 2001] defined impairment at the 5th percentile of normal and used statistical models to directly estimate BMDs and BMDLs corresponding to excess risks of 2%, 5%, and 10%. An excess impairment risk of 2%, or 20 per 1,000 children, would be considered unacceptable by most public health advocates. With continuous outcome measures for which impairment has been defined, the distribution-based BMD approach permits estimates of increased risk as a function of exposure (see Bailer et al. [1997]; Budtz-Jørgensen et al. [2001]; Crump [1995]). The criteria document reviewer refers extensively to the Environmental Protection Agency's (EPA) BMDS software (BMDS, Version 2.0, EPA 2009). This program, intended primarily for experimental animal data, does not have the capability of distribution-based BMD calculations following Crump et al. [1995] for continuous outcome measures that are based on epidemiological statistical models with covariates. For the diacetyl analysis of continuous pulmonary function measures, as pointed out by reviewers, there is abundant low dose (<LOD) observation. A substantial proportion (> 20%) of the population was exposed within a factor of 10 of the proposed REL. Furthermore, one would expect the low-dose behavior of diacetyl to be essentially linear because there are other environmental exposures already contributing to loss of pulmonary function (air pollution, smoking, second-hand tobacco smoke, etc.). Thus use of a linear statistical model was appropriate for the low dose region, and the model generally fit well (although not as well as the square root transformed cumulative exposure, possibly the result of variable susceptibility). Because statistical models implicitly extrapolate to the origin, and because we used a linear model, the result was identical to what would follow from specifying a point of departure and linear extrapolation. The proposed REL, at 0.005 ppm, represents risk to the overall population. Consideration of more susceptible groups, for example with refinement of the exposure-response model to nonsmokers, the 1/1,000 excess risk criterion would be associated with exposure levels below 0.001 ppm. On the other hand, consideration of the

		mild degree of impairment represented by the lower limit of normal would favor some adjustment to above 0.001 ppm.
5087	The history and regulation of substance vinyl chloride is a useful lesson to review. While it is recognized that diacytel and related compounds are not carcinogens, the disease that one might, develop can lead to one's death, and it makes little difference if one dies from a cancer or dies from respiratory insufficiency or the complications of a lung transplant. The commentary made by some that this is being regulated as if it was a carcinogen, instead of being inappropriate as suggested, is totally appropriate, given the potential loss of life that can occur from exposure. It should also be recognized that large amounts of diacytel and related compounds are currently used, and it would be reasonable to anticipate that if strict regulations were put in place for diacetyl and 2, 3-pentanedione that other compounds would be favored, and this criteria document should apply to them as well.	No response needed
5088	This first charge to a reviewer is whether the health hazard identification, risk estimation, and discussion of health *effects* was a reasonable reflection of the current understanding of the scientific literature. The simple answer is yes. Science is an ever moving adventure,and there is rarely, if ever, closure to scientific questions. NOISH has done an adequate job of identifying what is currently known and has used this information in an appropriate manner.	No response needed

	Philosophically, the approach taken by some who sent in comments would be to allow continued use until better data was available, inevitably leading to serious illness and death among workers until the facts are better proven to their satisfaction. This would be an unnecessary repeating of the vinyl chloride story where when disease was first indentified there had been prior information that had been ignored and regulatory activity had not been strictly enforced with regard to exposures that were thought safe. The important outcome from that regulatory activity was that workers no longer became ill after the imposition of the regulations regarding vinyl chloride after the hazard was identified.	
5089	The second charge is regarding additional critical studies relevant to occupational exposure. While there could always additional investigation of exposed populations, a more critical need appears to be appropriate laboratory methodically to identify levels in the workplace, and to create similar laboratory testing for related compounds. There might also be some benefit to more wide spread pulmonary function testing among exposed workers to better characterize the relationship between exposure to diacytel and related compounds and less severe changes in pulmonary function other than bronchiolitis obliterans and its severe respiratory compromise. It might also be also be useful to do some additional measurements of naturally occurring diacytel and related compounds and to see what levels actually occur in workplaces and cooking facilities when these compounds are liberated.	The reviewer makes some good points regarding research that would be beneficial. The authors agree on these points. Since the issuance of the review draft some work has become available regarding naturally occurring diacetyl and has been inserted into Tables 2.1 and 2.2. Also, additions to the criteria document regarding laboratory methodology have been made.
5090	The third charge has to do with recommended strategies for controlling and preventing exposures. These all seem reasonable and appropriate. There might be some leeway in allowing for personal protective equipment as long as it serves the same function and achieves the same outcome. It might not be necessary to require full-face respiratory protection if an alternate method,including the use of half-face with goggles would be utilized. One addition to this area would be the establishment of a registry for cases of bronchiolitis obliterans and other potentially severe pulmonary changes among workers exposed to diacytel and related compounds.	NIOSH policy is to recommend only full facepiece respirators when there is the potential for eye irritation. Half mask respirators with goggles are not being recommended because NIOSH is not aware of any standards for gastight goggles that would permit NIOSH to recommend such goggles as providing adequate eye protection. This policy is from the NIOSH Respirator Selection Logic [2004c] page 21.

5091	The fourth charge is asking if NOISH has a transparent and sound basis for setting a revised recommended exposure limit and I believe the simple answer to this is yes. It appears to be following all rules as required.	No response needed
5092	The fifth question is the quantitative risk assessment and if NIOSH should a ten-year duration instead of a 45 year duration. It is unclear why a presumed lifetime of work should not be looked at with only 10 years, perhaps, not being sufficient to see if disease develops.	Agreed. Both 10-yr and 45-yr estimates will be retained, and a 2.5-yr estimate has been added. Most of the cases did not arise in the short-duration/low cumulative exposure strata (Tables 5.16, 5.20. 5.22). We have repeated the analyses removing workers who had ever been mixers and see slightly *stronger* estimates of exposure response. This provides evidence that the effects are not largely arising from short-term high exposures.
5093	The sixth charge is any additional recommendations for worker protection and there are none at this time.	No response needed
5094	HAVING WORKER PROTECTION UNDERPINNINGS FOR A RECOMMENDED EXPOSURE LIMIT As noted above, there is a valuable lesson to be learned from the vinyl chloride story where the unwanted outcome of angiosarcomas of the liver were eliminated with a massive reduction in the exposure to vinyl chloride. It would be appropriate to lower as much as possible the exposure to diacytel and related compounds. There appears to no good basis for using the numbers as derived from smokers verses non-smokers, and the lower limit would protect both non-smokers as well as smokers. Also, there appears to be no good basis for the increased STEL for 2, 3-pentanedione following the same model used for diacytel and the STEL should be lowered to reflect that of diacytel. Any other compounds that are thought to be reasonably similar in their use or potential toxicity should follow these compounds in their setting of exposures limits as well. Limits should not necessarily be driven by what laboratory methodology is available to fix a current standard when this can be remedied in the future.	The current NIOSH [NIOSH 1995] REL policy states the following: "NIOSH RELs will be based upon risk evaluations using human or animal health effect data, and on an assessment of what levels can be *feasibly achieved* by engineering controls and measured by analytical techniques." Therefore, an analytical method must exist for the suggested REL and short-term exposure limit (STEL). The higher STEL for 2,3-pentanedione compared to the STEL for diacetyl reflects the analytical requirement of the 2,3-pentandione method.

5095	It must be recognized that regulatory bodies are required to make decisions in the face of imperfect science. This is clearly reflected in the many commentaries that were sent in regarding these materials. There are two approaches. One could allow continuing exposure and increasing numbers of injured and dead workers until better science is forthcoming, or one can take a more preemptive and precautionary approach and limit the exposures so that the majority of workers, if not all, will be protected and disease will be kept to a minimum. The idea that until further documentation and information is forthcoming, when there is little prospect of that occurring in a timely fashion, would be extremely detrimental to the health and well being of workers. What one sees in the written commentaries of some is the use of "sound science" and one has no difficulty in recognizing that the science upon which recommendations for exposure limits should be built should be sound. However, one must also recognize that there may be a great paucity of appropriate data,and that to wait for such documentation would put workers at risk. One finds particularly disturbing the argument that these materials are to be regulated as if they were carcinogens;this reviewer finds that entirely appropriate given that this is a potentially lethal exposure to some workers. If this was more of a reversible problem then one might take a different approach.	No response needed
5096	In summary,this reviewer finds the document appropriate given the current state of knowledge. The specific criticisms of areas of acute lack of information and standard setting are reflected above. The process by which NIOSH undertook a review seems entirely appropriate and in keep with current regulations. This is grave concern by this reviewer of the concept of 1in 1,000 being an inappropriate level for adverse outcomes in the workplace. Workers should not be thinking that they have this high of change of getting seriously ill, with probably some greater expectation than this rate for some potential lung damage although not as severe as the outcome of bronchiolitis obliterans. Given the imperfectness of science and the absolute need to regulate exposures and protect	See response to comment 5086.

	works with some recommendations as noted above this document appears appropriate.	
5097	Combined Exposure at Plants K and L:Plant G data excluded: There is always utility to grouping exposure response data and that it increases the potential power of the calculations and allows for tighter confidence limits. Specifically, for the consideration of Diacetyl at companies K and L, there is this utility in combining the data, but it should also be noted that the two plants in question are different kinds of facilities and may represent different types of exposures. That said, ultimately, the issue is not the nature of the plant but the ability of Diacetyl to cause illness and I would think that the combining and grouping the data would be useful in addressing this most important of many questions. It might be useful to present the data forK, Land G done individually, but then do a combination to see what the outcome is.	Data are presented for Plants K, L, and G. For risk assessment purposes, K and L have limitations as reported, and a REL derived on that basis would be much lower than 0.005 ppm. The draft criteria document addresses these issues, concluding that the heterogeneity across plants makes the pooling of Plant G with Plants K and L inappropriate. The reasons for preferring Plant G were stated in the draft criteria document and have been expanded on in the revision. Our current REL is based on the less restrictive definition in the case of the BMD analysis and on the more restrictive definition for the modeling of incidence and estimating excess lifetime risk. Both give similar results. A third, intermediate case definition has been added: FEV_1/FVC<LLofN. The high proportions of air samples < LOD at Plant K reflect the low exposures there (at time of survey) and were part of the reason to not use that dataset. Plant G has a wide range of exposures including many above 0.5 ppm. The G samples < LOD are not uninformative; in fact, they are the most precisely known exposures (when based on full-shift sample) and are vital for statistical modeling as they drive the intercept (baseline) estimate. The observed heterogeneity, a statistical observation, is one legitimate basis for deciding on which population to base a REL. If heterogeneity had been observed among otherwise equivalent populations (comparable retrospective exposure assessments, work history detail, etc.) then pooling would be appropriate, especially if the plants were thought to be a representative sample of the industry. Risk assessments for the pooled K and L plants are presented in the criteria document. See also responses to reviewers' comments 5114, EA-3, and 5130.

5098	Two Case Definitions: Since this document has as its goal the protection of workers the better approach, in my opinion,would be to use a less restrictive rather than more restrictive case definition. There are still many aspects of the disease entity that are not appreciated and if the goal is the protection of workers then a looser case definition with data addressing that would be preferable. As noted below, since the mechanism of this disease is not yet well appreciated, then a less restrictive definition would be useful. As more information becomes available over time it might then be appropriate to readdress this and see if a better understanding of mechanism and outcome then a more restrictive definition might be used. A value to using a less restrictive definition is that it would potentially catch more cases and ultimately be more protective of workers. It is always somewhat fraught with danger to use pulmonary function testing as a definitive outcome measure since it is so variable, both dependent upon confounding diseases or confounding factors such as smoking, or preexisting diseases such as asthma, and that it so individual dependent as to the validity of the pulmonary function testing.	See response to comment 5015.
5099	Preexisting Symptoms: As a clinician, it is not uncommon to have individuals not recall any prior associated symptoms until the clinical condition worsens to a point where the patient makes note of systematic changes. This is especially true in younger persons. Very often the change may be gradual, and suddenly the individual notes that they are, in fact, not feeling well. I think it is appropriate to use date of onset as a measure. Then one can characterize an individual whose spirometry findings qualify, if they have been vetted for prior exposure and if the findings are not easily explainable by a prior underlying condition or confounder. This would increase the number of individuals to be studied by a factor of two, given better quality outcome results. The risk one is taking is that excluding prior asymptomatic individuals biases data towards being less protective, rather that more protective. I feel it useful to include an alternate risk analysis including asymptomatic subjects who had abnormal pulmonary function test that are thought to be related to exposures.	This reviewer suggests that asymptomatic persons with abnormal pulmonary function not be excluded from the incident case analyses, and that the cumulative exposure be calculated to the time of abnormal pulmonary function test. In cross-sectional analyses of Facility G, one in four people with airways obstruction reported no respiratory symptoms. Approximately half (47%) of the entire cohort with abnormal pulmonary functions was asymptomatic; the suggestion seems to have merit in resulting in a more protective risk analysis. In response to this comment, p. 3-15 in Chapter 3 has been revised to include this additional information.

	Again, the guiding principle is the greater protection of the largest numbers of workers.	
5100	Cumulative Exposures Verses Employment Duration: It is clear from observing other settings with other types of exposures that employment duration may have very little to do with cumulative exposure. An example of how this is true comes from how cigarette smoking is often reported. The idea of pack years is, in fact, not very useful. A forty pack year history could be one pack per day for forty years, two packs per day for twenty years, or four packs per day one year, each having a different duration but some cumulative exposure. Duration is not a good measure, cumulative exposure is much to be preferred. A model that uses both is potentially a confounder; simple cumulative exposure (as best as can be determined) would seem to me to be best.	The cumulative exposure (e.g., pack-year) and duration assertions apply to the standard model where it is fundamentally assumed that susceptibility is unchanging over time; the surviving population is no different than the population at the start of exposure in terms of unknown confounders or innate exposure response. These conditions appear not to apply to this study population. So the conundrum is what to do in the absence of any marker for susceptibility? Duration in employment is a crude surrogate for susceptibility, longer duration being associated with lower susceptibility, but it appears to considerably improve model fit. It is not directly interpretable but provides a description of incidence that can be used in the risk assessment. In this context colinearity is not the issue; the effects estimates have opposite signs. The rate model produces a risk assessment that is close to the BMD approach. Modeling incidence without the duration term produces uninterpretable results. In the absence of markers for susceptibility or mechanistic sources of changing response, the only choice is a phenomenological modeling of the observed incident cases over time (duration). Table 5.11 pertains to the longitudinal analysis, which does not use duration. Asterisks were for footnotes, since replaced, and will be removed. See also response to comment 5135.

5101	Cumulative Exposure as a Quantitative Risk Metric: Part of the problem in answering this question is that the mechanism by which Diacytel causes disease has not been elucidated. To contemplate possible other agents and speculate as to mechanisms is of little use when it comes Diacetyl. If, in fact,the average exposure over one's working time has been seen to be a reasonable predictor of spirometry declines then it is reasonable to use that as a measure. The use of simple cumulative exposure,for reasons as explained in the question just above, may be less useful. It also may be that the average exposure is not the right metric and it might be more ideal, if the number can be determined, to see if peak exposures are helpful in determining who develops disease. might be that neither low level average exposure,nor cumulative exposure, is a particularly good predictor. In addition to looking at those metrics it might be well to consider if there is any correlation between peak exposure or highest exposure ever noted and clinical outcome.	See response to comment 5018.
5102	High Risk Subpopulation: Again, one is faced with the problem of not knowing what the mechanism of Diacetyl induced disease might be. If there is a somewhat idiosyncratic response among workers, or one that is dictated by genetic or indogenous basis, rather than simply exogenous exposure then modeling a subpopulation stratified by cumulative exposure and exposure duration may not be useful. It may also not be useful for the issues noted above. I feel that without a mechanistic understanding or the role of potentially short term high acute exposures being appreciated, then the method suggested is not entirely valid. That being the case, then data should be analyzed both ways with the caveat that neither may ultimately be seen to be correct.	The inclusion of a duration term into the modeling of incidence is necessary to interpret the results. In the absence of markers for susceptibility or mechanistic sources of changing response, the only choice is a phenomenological modeling of the observed incident cases over time (duration). See also response to comment 5019.

5103	Use of Other Models: Fully recognizing that there is extremely limited data by which to create a criteria document and generate a suggested level of exposure, it would be entirely inappropriate to use animal models other then as general guidance, and they should not, in my opinion, be utilized to establish human levels. While animal research, which I have done, has great utility for some things, there is enough variability from animal to humans with regard to some toxicological effects that this may not be an especially good utilization of an animal model to predict what will happen with human disease. What might be interesting is comparing what the human modeling shows us along with the animal data and if they are quite consistent, then in future one might consider using such animal data, with appropriate caveats, but if there is divergence I certainly would go with more limited human data and err on the side of caution rather than to with whatever the animal data might be telling us. Given many other animal models used to look at disease patterns, one knows that depending on what animal is used the variability and outcome can be great, and may not at all pattern what occurs in human populations. would therefore strongly urge that animal data not be utilized to create a standard to be used for humans and that the human data be utilized fully. As an adjunct, all modeling can be studied and one can see how close animal modeling comes to the human outcome, but one could conceive of a animal model giving a higher suggested value for exposure than the human data suggests, and would certainly be to the detrement of workers who are exposed.	No response needed
5104	Summary: These are the thoughts of this one peer reviewer and should there be any further clarifications needed I would be pleased to address then. It seems always appropriate to be conservative in setting exposure limits. If history teaches us anything we note that limits, over time, must often be lowered, sometime with great difficulty, and that the greater protection of workers will result from a conservative standard that could, with future additional data becoming available and mechanisms being better understood, be raised in workplace settings.	No response needed

5105	This review will focus on the following aspects of the draft criteria document: identification of and synthesis of relevant health literature; the transparency of NIOSH's review process; and, the quality of the risk assessment that forms the basis of recommended exposure levels.	No response needed
5106	Identification and synthesis of relevant literature: It appears that NIOSH has identified the most relevant literature for developing a proposed exposure limits. Clearly, the NIOSH study (Kreiss 2002) that originated as a series of health hazard evaluations is pertinent, and has arguably assembled the most extensive data for determining exposure-response relations for diacetyl (DA) and pulmonary impairment. Descriptions of the other studies in the US (Lockey 2009) and the Netherlands (van Rooy 2007, 2009) are relatively perfunctory, which is unfortunate insofar as data from these studies, in theory, might be considered for comparative risk assessments (see below). Overall, the review of literature was thorough, especially for the NIOSH studies. However, far too much detail of the research was presented; a more succinct summary would be easier for the reader to assimilate. Also, the literature review was largely uncritical, as specific study strengths and limitations were not clearly articulated. Exposure assessment was discussed in some detail, but there is little discussion about potential selection bias and confounding.	Various reviewers have commented that the document contains too much or too little detail. This comment suggests the document overly describes the studies in question but does not include a critical review of those studies. While some minor modifications have been made in the descriptions of the studies reviewed by this document, the data presented in those studies has been accepted and used where applicable in the evaluation of exposures and risk for diacetyl and 2,3-pentanedione. See also responses to comments 5113 and 5114.
5107	Transparency of the review process: The process by which NIOSH identified literature and performed risk assessments seems suitably transparent. The document is very explicit about which studies were identified and the health endpoints that were considered ultimately for risk assessment.	No response needed
5108	Quality of the risk assessment: The choice of reduced FEV1 as the primary risk assessment endpoint is well justified. As the authors of the document note, reduced lung function is a major component of the index condition, bronchiolitis obliterans (BO), and FEV1 impairment at levels below those leading to clinical diagnoses of BO have been observed in exposed workforces. Risk assessments for percent predicted FEV1 and frequency of cases of FEV1 or FEV1/FVC below the lower limits of normal are readily interpretable clinically and epidemiologically. A particularly challenging aspect of the risk	Risk assessment for 2–5 yrs: This is an appropriate proposal in view of the apparent early effects that we have interpreted as high susceptibility. The suggestion does not necessarily imply that lifetime risk should be estimated based on 2–5 yr exposure and could be interpreted to suggest that a 45-yr exposure be treated as a composite of multiple 2–5 yr exposures (corresponding to jobs held by 22 (for 2 yr) to 9 (for 5 yr) workers). The 2–5 yr risk assessment has been added to the document using both the benchmark dose and rate modeling approaches.

	assessment was taking into account the time course of pulmonary impairment. Evidently, diacetyl can induce severe acute lung injury within a short time period (within months of exposure). As such, the standard approach of risk assessment for 45 years workplace exposure may not be especially meaningful, particularly when workers exposed to diacetyl typically work for relatively short durations (<1-5 years). The authors have presented risk assessments for 10-year exposure. It might also be worthwhile to preforming risk assessments for shorter exposure durations (e.g., 2 or 5 years).	
5109	Some of the arguments advanced in this section of the document appear to be speculative, or at a minimum, deserve further explanation. In particular, the cumulative dose metrics of the summation of the square root (or square) of exposure intensity, described on p119, are not standard for risk assessment. As the authors note, simple cumulative exposure is the typical default metric. In fact, the explanation of interpretations of these alternative metrics seems counterintuitive. Likewise, the concept of "susceptible" subgroups is very difficult to follow, and has the appearance of speculation. Perhaps some simple examples (e.g., in an appendix) would help readers who had similar difficulties as this reviewer.	The examination of dose-rate using square root and squared intensity is not standard practice although given the concern for special risk in mixers, it is appropriate to test for a dose-rate effect, which the cumulative square root (or square) exposure metrics do. The variable susceptibility issue has been re-examined, and a new, more traditional interpretation has been presented referencing a survival effect.
5110	As alluded to previously, two other studies of diacetyl-exposed workers have been conducted, and their results published (Lockey 2009, van Rooy 2009). Both studies generated relevant clinical and exposure data, enabling dose-response estimation. Therefore, the document authors might consider requesting access to data from these studies for the purpose of performing replication risk assessment. Similarity of risk assessment findings among the three databases would greatly strengthen the conclusions in the present document.	Considerable effort was put into an attempt to utilize the van Rooy and Lockey data [Lockey et al. 2009; van Rooy et al. 2009]. The van Rooy dataset is smaller but exhibits the same pattern as the NIOSH index facility G study data [NIOSH 2006] and possibly more extreme, showing evidence of a strong healthy worker survivor effect. With increasing cumulative exposure, the average FEV_1 also increases. If duration of exposure were included in a model, we predict that FEV_1 would increase with duration while diminishing with cumulative exposure. The analysis in van Rooy et al. [2009] did not accommodate a susceptibility hypothesis, although presumably that could be done by NIOSH using the van Rooy [2009] data except that, unlike NIOSH index facility G [NIOSH 2006], the van Rooy [2009] dataset does not have symptom onset information. The Lockey dataset would have the same limitations that led NIOSH not to use it: poorly defined exposures prior to the health hazard

		evaluation. If the analysis were limited to new hires or a longitudinal analysis since the health hazard evaluation then that would not be a problem, but exposures after 2005 were low and statistical power would probably be limited. The mean duration in production for the cross-sectional population identified 2005–2006 in the Lockey [2009] dataset was reported to be ~7 yr, implying that many (possibly 300) were new hires since the health hazard evaluation exposure assessment in 2003. On the other hand, exposures since the health hazard evaluation were quite low and so any analysis, including one addressing possible susceptibility issues, would be quite limited. See also response to reviewers' comments 5135 and 5114.
5111	Finally, the risk assessment chapter, although technically impressive, is no doubt difficult to follow for readers with little background in risk assessment methods. Inclusion of some lay language explanations, perhaps as brief sections at the beginning (preface) and end (summary) of the chapter, would be welcome additions.	Sufficient lay language has been added to the executive summary.
5112	This review consists of comments offered in response to questions posed in the original charge, followed by responses to additional questions raised by OSHA and NIOSH personnel. Please note that I will not attempt to provide responses to questions that are outside my area of expertise.	No response needed
5113	The document is very explicit about which studies were identified and the health endpoints that were considered ultimately for risk assessment. It appears that NIOSH has identified the most relevant literature for developing a proposed exposure limits. Clearly, the NIOSH study (Kreiss 2002) that originated as a series of health hazard evaluations is pertinent, and has arguably assembled the most extensive data for determining exposure-response relations for diacetyl and pulmonary impairment. Descriptions of the other studies in the US (Lockey 2009) and the Netherlands (van Rooy 2007, 2009) are relatively perfunctory, which is unfortunate insofar as data from these studies, in theory, might be considered for comparative risk assessments (see below). Overall, the review of literature was thorough in terms of identifying the most relevant health effects, especially identified in the NIOSH studies. The document would,	We appreciate that the reviewer considered the review of literature as thorough. The reviewer suggests that a less detailed review would benefit some reviewers. He also comments that the review is “largely uncritical.” This is by design, as are the presentations he notes on selection bias and respiratory endpoint selection.

	however, benefit from improvements in presentation. In particular, far too much detail of the research was presented; a more succinct summary would be easier for the reader to assimilate. Also, the literature review was largely uncritical, as specific study strengths and limitations were not clearly articulated. Exposure assessment was discussed in some detail, but there is little discussion about potential selection bias and confounding. The choice of reduced FEV1 as the primary risk assessment endpoint is well justified. As the authors of the document note, reduced lung function is a major component of the index condition, bronchiolitis obliterans (BO), and FEV1 impairment at levels below those leading to clinical diagnoses of BO have been observed in exposed workforces. Risk assessments for percent predicted FEV1 and frequency of cases of FEV1 or FEV1/FVC below the lower limits of normal are readily interpretable clinically and epidemiologically.	
5114	As mentioned above, two studies from, respectively, the Netherlands (van Rooy 2007, 2009) and the US (Lockey 2009) deserve more intensive review, similar to what was performed for the NIOSH study (Kreiss 2002). Moreover, both the van Rooy and Lockey studies generated relevant clinical and exposure data, potentially enabling dose-response estimation. Therefore, the document authors might consider requesting access to data from these studies for the purpose of performing replication risk assessments. Similarity of risk assessment findings among the three databases would greatly strengthen the conclusions in the present document.	In 12 or more instances the studies of van Rooy and Lockey were discussed by peer reviewers and in 10 instances by public commenters. The van Rooy dataset is smaller but exhibits the same pattern as the NIOSH index facility G data [NIOSH 2006] and is possibly more extreme, showing increasing FEV_1 with cumulative exposure. Applying the traditional interpretation would lead to the conclusion that diacetyl exposure is protective. The analysis in van Rooy et al. did not accommodate an employee selection/survival or susceptibility hypothesis although there is evidence for it. Unlike the NIOSH index facility G study [NIOSH 2006], the van Rooy dataset does not have symptom onset information so that incidence cannot be modeled. The original Lockey dataset has unknown exposures prior to the health hazard evaluation in 2003. Three additional ConAgra plants studied by Lockey and White et al. have no exposure data prior to 2005. If an analysis were limited to new hires with multiple evaluations since the first survey, or if a longitudinal analysis were performed, then exposure-associated trends could be estimated, but exposures after 2003 or 2005 were quite low. Small numbers of qualifying subjects and low exposures would mean very limited statistical power. Like the van Rooy study, the published Lockey analysis made no accounting for employee

		selection/survival or susceptibility. Several of the comments regarding the Lockey data also question the appropriateness of the study criteria and obviously the conclusion drawn from the data. Many of those suggest an increased reliance on the Lockey data or collection of additional data to supplement that information. See also response to comment 5135.
5115	Exposure control and medical surveillance are outside of my area of expertise.	No response needed
5116	The process by which NIOSH identified literature and performed risk assessments seems suitably transparent. Likewise, the risk assessment methodology and approach to recommending exposure limits was sufficiently transparent insofar as the methods are described in detail. However, the risk assessment chapter, although technically impressive, is no doubt difficult to follow for readers with little background in risk assessment methods. Inclusion of some lay language explanations, perhaps as brief sections at the beginning (preface) and end (summary) of the chapter, would be welcome additions.	Sufficient lay language has been added to the executive summary.
5117	A particularly challenging aspect of the risk assessment was taking into account the time course of pulmonary impairment. Evidently, diacetyl can induce severe acute lung injury within a short time period (within months of exposure). As such, the standard approach of risk assessment for 45 years workplace exposure may not be especially meaningful, particularly when workers exposed to diacetyl typically work for relatively short durations (<1-5 years). The authors have presented risk assessments for 10-year exposure. It might also be worthwhile to preforming risk assessments for shorter exposure durations (e.g., 2 or 5 years).	See response to comment 5092.
5118	None that I can think of, although exposure control is not my area of expertise.	No response needed

5119	In general, exposure-response analyses that form the basis of risk assessments should include as much valid exposure and health outcome data as can be assembled, provided that the data were obtained in unbiased and relatively similar manner across subsets (in this case, plants). Combining datasets maximizes both the likelihood of estimating effects across a broad exposure range, as well as statistical power. In the current situation, combining plants K and L seems appropriate, yet a case could be made for including plant G (at least the exposure and health outcome data that are similar in nature). Thus, separate exposure-response modeling could be performed for: a) plants K and L combined; plants K, L, and G combined; c) separately for plant G. Similarity or differences among exposure-response models would indicate the robustness of the findings to differences in study populations and exposure circumstances.	The draft document addresses these issues, concluding that differences make the pooling of Plant G with Plants K and L inappropriate. See response to comment 5097.
5120	This is not a topic of my expertise.	No response needed
5121	The clinical aspects of this issue are outside of my expertise. However, in terms of general epidemiological considerations, it would seem that if spirometric abnormality is the outcome of interest, then presence or absence (or recall) of symptoms is irrelevant, i.e., asymptomatic workers should be included. (Note that I will defer to responses based on clinical grounds which might override epidemiological concerns.)	See response to comment 5016.
5122	Use of employment duration as a covariate in the analysis of associations with cumulative exposure is a poor choice for the stated reason of collinearity. More fundamentally, duration is a component of cumulative duration (duration x average exposure intensity = cumulative exposure); thus, including both in the same model is not logical. Analyses of duration, adjusted for intensity, and vice versa, are valid, and can be informative in some situations.	While these assertions apply to the standard model where it is fundamentally assumed that susceptibility is unchanging over time (e.g., the surviving population is no different than the population at the start of exposure in terms of unknown confounders or innate exposure response), these conditions appear not to apply to this study population. See also response to comment 5017.

5123	Cumulative exposure is typically the default metric in occupational and environmental epidemiology because it is a proxy for true dose. For health outcomes that result from accumulated biological damage, cumulative exposure is the obvious choice. Some biological effects are due to above-threshold "peak" exposures. However, defining peak exposure can be problematic and subject to erroneous assumptions and biases. Average exposure is sometimes applied in exposure-response analyses and risk assessment to estimate dose rate effects. There are, however, some severe limitations to average exposure as a metric. First, with the exception of an unvarying exposure intensity throughout employment (which is often not a realistic scenario), average exposure cannot be linked with specific time periods in life or with specific biological processes. Also, exposure variability, as would be of interest to detect effects of peaks, cannot be discerned from average intensity. As such, average intensity has less interpretability than cumulative exposure, peak exposure, or exposure duration, in an epidemiological or risk assessment context. Average exposure might, however, be applied as a very imperfect proxy for peak exposure when time-specific exposure data are fragmentary.	Agreed. See response to comment 5018.
5124	As presented in the document, the concept of "susceptible" subgroups is very difficult to follow, and has the appearance of speculation. Perhaps some simple examples (e.g., in an appendix) would help readers who had similar difficulties as this reviewer. This concept seems too abstract to evaluate potential consequences on the risk assessment modeling.	An attempt will be made to better justify the interpretation of susceptibility. See also response to comment 5019.
5125	Toxicology and applications of animal toxicology data to risk assessment are outside of my area of expertise. Nonetheless, I will offer the opinion that, where available, sufficiently valid and complete epidemiological data should be given higher priority than animal toxicology data for risk assessment purposes. Risks assessments based on animal toxicology data may serve as indirect replications of those based on epidemiological data, although divergence of results should not detract from confidence placed on the epidemiological data. (NB: Please consider the source of this comment—from an epidemiologist!)	We agree with the reviewers comment.

5126(a)	NIOSH's Criteria Document for recommended exposure limit for occupational exposure to diacetyl and 2,3-pentanedione (thereafter *Document*) is well structured and for most part well presented as well. It contains a comprehensive review of the published literature and available data concerning exposure to diacetyl and 2,3-pentanedione, its associated health risk for exposed workers, and toxicity experiments using animals. NIOSH selects a few studies and datasets to conduct exposure-response assessment, and from which derives risk estimates and exposure limits. While the information presented in the Document reflects the state of the science to a good degree, it does NOT clearly outline the strategy with which the literature search and review were conducted (e.g. publication period that was covered, types of journals or electronic databases searched, whether non-peer reviewed reports included). It also does not state inclusion and exclusion criteria, if any, for selecting studies to review and to advance for use in the formal risk assessment. The absence of such inclusion and exclusion criteria may have the following consequences for the Document: (1) reduced transparency; (2) compromised consistency in synthesizing information arising from difference sources and of varying quality; (3) non-disclosed uncertainty and variability in risk assessment. To gather and synthesize scientific information across multiple disciplines, systematic review has become a principle and common standard for practice in medicine and health as well as in other fields. (The Campbell Collaboration and Cochran Collaboration are two main sources of activities, methodologies, and results of systematic review.) The Document can greatly benefit from and be strengthened by adopting the approach of systematic review, an argument made by several NAS reports concerning EPA's IRIS documents. NIOSH has undoubtedly conducted a comprehensive review of the scientific literature, but the process can be made more transparent and rigorous if guided by the principles of systematic review.	As requested, we have added a description of how literature searches were conducted in Chapter 1, section 1.2 scope.

5126(b)	For the Document, a systematic review may include: 1) A well-defined literature search strategy. For example, consider a cut-off point in time up to which all publications will be searched. What electronic databases should be searched? Should non-peer-reviewed publications such as government reports be included? 2) Well-defined inclusion and exclusion criteria for selection of studies/datasets to advance to the next stage(s) of analysis. Important issues include if adequate exposure data are available, relevance of the health outcomes, presence of exposure-response relationship in the data, sample size of the study, study design (case reports vs. epidemiological studies). A consistent data analysis plan for the epidemiological study data would also strengthen the Document. The Document considers 2,3-pentanedione similar to diacetyl in its health risk to human largely based on similar chemical structure. Except for an animal pilot experiment, most of the studies and data reviewed were on diacetyl. Are there additional evidences that support this fundamentally important assumption of toxicological similarity?	While some reviewers consider this a good or maybe overly complex literature review (e.g., comment 5113), this comment asks about search criteria. An addition was made to Chapter 1, section 1.2, which responds to the general question asked. Further, the criteria mentioned in this comment have been considered by the document authors, although they were not included in the document. The analysis plan for the epidemiological study data is discussed in some detail in those chapters. And the inclusion of 2,3-pentanedione in the document was based on available information and given significant consideration.
5127	While there does not appear to any additional critical studies to be included, the treatment of the studies reviewed in the Document does not appear to be even. The Document reviews a number of NIOSH HHE studies including the cross-sectional medical and environmental surveys at six microwave popcorn factories (including the index plant) as well as those conducted by private investigators (Lockey et al, 2009), the California Department of Public Health lung disease prevention program for workers of California flavoring manufacturing plants (Kim et al, 2010), and the Wisconsin flavoring manufacturer (NIOSH, 2009d). For quantitative risk assessment (e.g. exposure-response assessment), however, the document relies eventually on the data from index popcorn factory. It stopped advancing other studies long the formal risk assessment process at some point, but does not clearly state the inclusion and exclusion criteria for such a choice/decision. If a set of inclusion-exclusion criteria were in place, it would not be hard to justify why some studies were not advanced. For example, the California studies (Kim et al 2010; Kreiss et al. 2011) do not have adequate exposure data	The inclusion criteria for studies serving as the basis of the risk assessment were adequate numbers of employees, representative diacetyl exposure data, and the existence of a job-exposure matrix (longitudinal if exposures changed and serial lung function data were available). For health outcomes such as pulmonary function decline (continuous) or abnormal pulmonary function decline (categorical), serial data were required. Four microwave popcorn facilities (G, K, L, and N) had adequate numbers of employees, of which only Facility G had serial exposure and spirometry data. No flavoring facilities that NIOSH studied had adequate diacetyl exposure data. These inclusion and exclusion criteria are discussed in the risk assessment chapter. Dr. Lockey's longitudinal exposure and spirometry data through 2012 were an alternative possibility. However, these data have not been made available to NIOSH for risk assessment and had no historical exposure data for three of the plants; in the fourth plant, NIOSH measurements in 2003 were likely underestimates because changes were made immediately

	for exposure-response assessment; adequate exposure-response data can be a selection criterion. Lockey et al (2009) and White et al (2010) reported a cross-sectional survey by a large food company at four microwave popcorn factories. The study provides job-exposure metric and adequate health effects data (spirometry). Although the authors did not find statistically significant association between exposure to diacetyl and rapid lung function decline based on three spirometry tests six-month apart, they did report significant association between dichotomized exposure level and lung function decline. The Document failed to state clearly the reasons this study was excluded from further analyses. Is it because the authors did not conduct dose-response analysis that can directly be used to support risk assessment? Is it because a re-analysis of the original data is required but the original data are not available?	before the NIOSH field study, including isolation of the mixing room and addition of exhaust ventilation.
5128	The Indiana study (NIOSH, 2008b, 2011b) appears to have both exposure and lung function decline data. (But it was not summarized in Table 3.1.) The *Document* needs to explain why this study is not advanced to the next level of analysis. Similarly, the Document could have explained the reasons that the diacetyl manufacturing study by van Rooy et al (2007; 2009) was not advanced. Was it because of limited gradation of exposure information or lack of exposure-response relationship? More critically, the Document needs to clearly state why only data from factory G (the index factory), K and L were considered for dose-response modeling, and the other three factories in the NIH HEE surveys were excluded. While NIOSH is most probably able to justify its inclusion and exclusion, failure to have done so in the Document creates the perception of "cherry picking" and bias. Having a set of well-defined inclusion-exclusion criteria at each stage of a risk assessment can greatly strengthen the risk assessment process and improve the transparency and consistency of the Document.	We have added the NIOSH 2011b study to Table 3.1. The historical exposure data available from this company consisted of nine personal samples for diacetyl (in four locations) between March 2004 and July 2007. Five of these were below the minimum detectable concentration using NIOSH Method 2557, without information on temperature and relative humidity necessary to correct these underestimates. NIOSH found diacetyl in thermal desorption tubes in two of four locations in 2008, but these were not quantitative measurements. The company supplied 45 personal sample measurements in eight locations in 2008–2009, of which 14 were below the minimum detectable concentration using OSHA Methods PV2118 and 1012. These 35 measurements were considered insufficient for quantitative risk assessment. The exposure information is now included in Table 2.1**.** The van Rooy et al. papers [van Rooy et al. 2007; van Rooy et al. 2009] reported that 26 area samples (82–219 minutes) and four personal task-based samples (33–90 minutes) were taken between 1995 and 2003. Many jobs were not sampled. These data were insufficient for quantitative risk assessment over the period of plant operation from 1960 to 2003. This information is now included in the description of the literature in Chapter 3.

5129	The preventive measures discussed in the Document make sense. The effectiveness of engineering controls and work practices (increased isolation of and improved ventilation in mixing rooms, and use of personal protective devices) in reducing workplace concentration/exposure is well documented (e.g. in index popcorn company). The subsequent reduction in health risk is anticipated, and is also documented to a good degree. Medical surveillance system is highly valuable and will provide useful data for future risk/safety evaluation. The analyses of eight NIOSH medical surveys (Kanwal et al 2011) to compare the health effects before and after the implementation of engineering controls in the index plant show unequivocal evidence that lung function decline slowed to what is expected of a normal population two years after the engineering controls being implemented. However, the comparison between the two exposure groups is confounded by the distinct exposure history. (Group 1 consists of workers hired before the engineering control who had longer employment history and group 2 consists of people hired after the engineering control who had an average six-month employment history).	The reviewer notes that the group of workers present at microwave popcorn Facility G in November 2000 differed from the group of workers hired after interventions began (2001–2003), with the latter group having an average 6 months tenure in comparison to an average tenure of about 6 years in the former group. Cumulative exposure to diacetyl in the two groups differs not only because of vastly different tenures, but also because of the decrease in average exposure levels for the period 2001–2003 for both groups. As the reviewer notes, the differences in symptom prevalence in the two groups is marked, as are average pulmonary function indices. The data that support normalization of lung function decline is based only on persons who participated in all eight cross-sectional surveys, so that result is not dependent on characteristics of the newly hired workers in 2001–2003 [Kreiss 2007]. Average exposure of the two groups over their differing tenure differed, and this characteristic parallels the change in symptom prevalence and pulmonary function. The only way to know that the newly hired workers remain healthy in the succeeding years would be to have long-term follow-up. This information is not available. No changes have been made to the document in response to this comment.
5130	The overall approach of NIOSH in revising the REL is sound. But at places the Document suffers from a lack of transparency and sound justifications, due largely to inadequate technical presentation or approaches. Chapter 5 is of particular concern where some parts of the presentation have created unnecessary confusion about NIOSH's approach in deriving the risk estimates. It should be revised to be more relevant, informative, and readable. The Document presents large amount information on the exposure-response modeling process, including trials for many exposure metrics, different outcomes, various regression models for several continuous outcomes, Poisson regression models for lung function impairment incidence, and dichotomizing schemes of continuous outcomes ppFEV1 and FEV1/FVC. This effort is highly commendable. Unfortunately the distinction among these different modeling options is somewhat lost in terms of the purposes and technical strength and weakness of each (albeit there is a discussion of	Chapter 5 has been revised to address these and other comments, and tables have been improved, to make the content more easily understood. The progression from statistical models to risk assessment procedures has been made more explicit, and the individual steps in the two procedures are presented in more detail. Five peer review comments questioned the use of data from one plant to define the exposed population. The analyses upon which the risk assessment is based were entirely focused on diacetyl vapor exposures. There were no useful, comprehensive air sampling data for diacetyl in any other form in populations with respiratory outcomes measured, and the historical contributions of powdered or encapsulated forms in the studied workplaces is unknown; diacetyl exposure in particulate form has not been investigated in any detail. Whether particulate forms are less (or more) toxic is not known but a valid question for future research. The choice of plants to analyze for the risk assessment was entirely

	sensitivity in section 5.5). It would be preferable that the Document highlight the final models used for risk assessment (e.g. benchmark dose estimation) vs. models for other supportive purposes, and discuss how these models are derived and how they would be used. It would also be helpful to distinguish the different approaches/models MIOSH used for risk estimation. The following comments are for specific subsections.	determined by the quality of information available from many candidate sites. The number of plants selected has little inherent significance for generalizability on the effects of diacetyl as vapor, provided that is the dominant form of exposure. Generally more plants implies greater statistical power but with the possible cost of more exposure misclassification due to diminishing quality of retrospective exposure assessments or work history detail. The final decision on using the NIOSH index facility G [NIOSH 2006] was based on data quality but also on (a) the relative confidence that exposure levels prior to the NIOSH index facility G [NIOSH 2006] health hazard evaluation had not changed materially, and (b) the observed heterogeneity of exposure response at two other candidate plants at least one of which had important exposure history lacking.
5131	Section 5.1.1: Out of six microwave popcorn factories involved in the NIOSH HHEs, two have already been dropped without any justification. NIOSH judges the remaining four to have potential for use in risk assessment, but quietly dropped factory N from risk assessment beyond the point of a discussion of airborne concentration in the factory.	See response to comment 5130.
5132	Sections 5.2.1 and 5.3.1: There appears to be an excessive discussion on exposure metrics that do not have adequate biological and pharmacokinetic underpinnings. These exposure metrics involve "mathematical transformation" - squared and square-root of daily average or cumulative concentrations. There is a considerable risk of choosing among these mathematically-oriented metrics on the basis of non-significant and negligible changes in the data variation (as measured by R^2) captured by a model. This appears to be the case in section 5.3.1 where the discussion of "best" or "stronger" predictors does not seem to have any biological justification or statistical significance. Caution should be exercised in using R^2 to judge a model. R^2 is only one of many measures of model adequacy and it depends heavily on intrinsic data variation. As a result, the discussion of "better" predictor on page 123 has only amplified inconsistency among the exposure metrics as different outcomes would suggest different metrics. For instance (P123, line 5-6) the	The metrics examined address biological issues such as selection confounding and survivor effects, i.e., the changing composition of unmeasured population risk factors over time or over exposure experience. R^2 is, as the reviewer implies, a blunt tool for choosing the preferred model, particularly with higher correlated metrics. Other factors contributed to the decisions, as described in the criteria document. Dose-rate is also a biologically grounded issue as evidenced by the considerable interest in peak exposure effects. Cumulative square root or square concentrations is a limited but useful approach for assessing dose-rate effects with the available exposure data. The explanation and interpretation of the dose-rate findings have been expanded to address these points. Seven or more comments addressed the dose-rate analyses. Because cases of bronchiolitis obliterans were initially observed in mixers, the popcorn workers with intermittent but highest exposures, the hypothesis naturally followed that peak exposures or only the

	interpretation of cumulative square root metric is awkward because of the lack of meaningful interpretation. The resultant exposure-response relationship is that of supra-linear, and the word "negative" is both ambiguous and misleading. The possibility of a high-risk sub-population susceptible to earlier onset under lower exposure, as suggested by the Document, explains in large part why a supra-linear exposure-response model, hence the artificially good fit of cum($DA^{1/2}$), was observed. The discussion on these mathematical metrics can be substantially cut.	highest exposures were placing workers at risk. No exposure assessment was available for a full worker population that would permit analyses of the contribution of peak, intermittent exposure; however, calculation of cumulative exposures using various powers of the estimated air concentrations would permit testing whether the risks increased more or less than proportionately with increasing average (8-hour [hr]) exposure level, when summed over time. In the NIOSH index facility G analysis [NIOSH 2006], somewhat better model fit was obtained using the metric, cumulative square root of concentration, arguing against the hypothesis that risk is limited to higher exposures, but this observation may be confounded by the susceptibility issue. A more convincing test resulted from analyses which excluded workers who had ever been mixers; in these analyses the exposure response parameter estimate (the slope of the exposure response) was slightly larger (this is now reported in criteria document revision).
5133	Sections 5.2.2 and 5.2.3: NIOSH conducted regression models for continuous outcomes ppFEV1 and FEV1/FVC, longitudinal regression models for repeated ppFEV1 and FEV1/FVC of workers at the index factory. It also conducted Poisson regression models given person-year exposure for the incidence of pulmonary obstruction defined by FEV1<LLN (5^{th}-percentile of population) or FEV1<LLN & FEV1/FVC <LLN. Company L & K were pooled for combined analysis. (See also responses to Additional Question #1). While such modeling efforts are valuable, it is not easy to see the purposes and objectives of these models. Some outline for each and every model with respect to its objective and purpose would be helpful. It is especially helpful to point out the models selected for risk estimation.	See response to comment 5130.

5134	Section 5.3.1: The incorporation of pack-years squared in the regression model for ppFEV1 of plant G workers is not apparent in reason (Table 5.6). The protective impact of "ever smoking" on ppFEV1 invites scrutiny. Take the first model for example (Table 5.6). Only when a person smoked more than 12 pack-years than smoking started to impact lung function decline; those never smoked would have predicted ppFEV1 6.7% lower than those smoked 1 pack-year. In contrast, such an "ever-smoking" effect is not present (non-significant) in the Poisson regression models for lung obstruction incidence data (Tables 5.13 and 5.15); nor is it present in the Poisson regression model with the relative incidence rate of a high-risk subpopulation. The ever-smoking effects hardly make any sense and it appears to be a statistical artifact of data variation. Its carry-over effects on risk estimation are also of serious concern. It is also puzzling why "age" is not a confounder retained in the models. The Document states the company N produced regression results similar to company G. This appears to be the last reference to company N. No explanation is given as to why company N was removed from any further analysis.	Chapter 5 has been revised to address this and other comments, and tables have been revised to make the content more easily understood. The use of pack-yrs squared permits a more general fit, accommodating survivor effects. The ever smoker term does invite scrutiny but is highly significant, suggesting the selection process proposed. Its absence in the rate model reflects that the rate of new cases (rate of change of FEV_1, etc.) doesn't depend on ever-smoker status. Age is adjusted for in the models with percent predicted and thus not included as terms, but is included in the incidence rate models as a potentially important confounder. Plant N had one breathing zone air sample for mixers and was a very small population, about 1/10th the size of the Plant G population. Early analyses with Plant N were consistent with Plant G but lacking statistical significance. Excluding it or pooling it with Plant G was a decision of no consequence. Two other plants where health hazard evaluations had been performed had absent or insufficient work history or exposure assessment for any meaningful analysis [NIOSH 2003b]. The progression from statistical models to risk assessment procedures has been made more explicit, and the individual steps in the two procedures will be presented in more detail.

5135	Section 5.3.3: The negative duration term in the models for the incidence of lung obstruction is problematic. The inclusion of the quadratic terms $(age\text{-}40)^2$ and $packyr^2$ in the models is puzzling and needs justification. See response to additional questions #4 and #6 as well.	The quadratic terms merely relax some of the model requirements to permit more flexible relationships. The only downside would be loss of statistical power. Nine or more peer reviewers' comments related to the interpretation of negative duration effects. In the standard modeling of incidence rate it is fundamentally assumed that inherent susceptibility is unchanging over time: the surviving population, when adjusted for age and accounting for interactions, is no different than the population at the start of exposure in terms of unknown confounders or innate exposure response. The experience of the microwave popcorn workers studied suggests that this assumption does not apply to this population. The statistically quite significant but paradoxically opposing joint effects of exposure duration and the diacetyl exposure metrics (for which estimated parameters have opposite signs) suggests that underlying assumptions have been violated. Usually, exposure duration is positively associated with cumulative exposure, and including both terms would tend to diminish the strength of association for each. With diacetyl, including both terms dramatically increases the strengths of association. If susceptibility is declining with employment duration, the conundrum is what to do in the absence of any marker for susceptibility. Duration in employment is a crude surrogate for susceptibility, longer duration being associated with lower susceptibility, but it appears to considerably improve model fit. It is not directly interpretable but including duration in the model provides a description of incidence that can be applied in the risk assessment. Modeling incidence without the duration term produces uninterpretable results. In the absence of markers for susceptibility or mechanistic sources of changing response, the only choice is a phenomenological modeling of the observed incident cases over time (duration). Besides observing this phenomenon (in the present analysis) in the NIOSH index facility G population [NIOSH 2006] and pooled ConAgra and American Popcorn populations [NIOSH 2004a, b], it was also probably present in the Dutch study by van Rooy et al. [2009] where FEV_1 was observed to increase with cumulative exposure in a group of chemical manufacturing process workers (n=95) with quite

		high levels of diacetyl over a 40-year period and with several cases of bronchiolitis obliterans. A plausible interpretation here is that workers experiencing adverse respiratory effects at the low end of the FEV_1 distribution are leaving employment sooner than others, thereby raising the average FEV_1 in the remaining population. Additionally, if there are individuals with below-average FEV_1 and a higher susceptibility (losing breathing capacity faster than others) this effect would be accentuated. Duration of employment now becomes a dependent variable, determined in part by exposure levels.
5136	Section 5.3.5: The interpretation of the models should change to reflect the recommended changes to be made to the previous sections.	The interpretations now presented in the revised criteria document reflect the current results, which include several new analyses and a modified understanding of the susceptibility issue.
5137	Section 5.4.1: This section is considerably inadequate, and should be carefully reviewed and substantially revised. For example, it would be helpful to include a brief and concise description of the benchmark dose methodology in section 5.4.1.1. It would be particularly helpful to explicitly point out how a chosen exposure-response (regression) model is converted to risk (prevalence) of a defined impairment, and how the prevalence is then used to define a benchmark dose. Using mathematical notations and formula would be appropriate.	See response to comment 5130.

5138	Section 5.4.1.2 begins by defining the prevalence (probabilistic risk) of abnormal FEV1, (a) Prob(FEV1<5th-percentile|exposure) or (b) Prob(FEV1<60% Predicted Value|exposure) based on a regression model (model 1 in Table 5.6) of FEV1 for data of the index company G. The benchmark dose is based on excess risk due to exposure over the spontaneous risk of a reference population (e.g. exposure=0), which is also estimated from the regression model. Although the BMDs (Table 5.27) based on (a) are universally applicable to all subgroups as characterized by covariates (e.g. gender, smoking) because of the linearity of the regression model, those based on (b) are in principle different across subgroups. The Document does not discuss this important distinction. Moreover, if interactions between exposure and other factors (e.g. exposure impacts men and women differently) are possible, the BMDs in both cases would vary across subgroups. The Document does not discuss this issue either. The Document then considers an alternative approach in which NHANSE III data were used to estimate the prevalence of abnormal FEV1 of the control population. This, I think, involves the determination of LLN, the prevalence of ppFEV1<LNN, for example, for each subgroup in the NHANSE population using the regression model. The process is unclear in part because the Document (page 130, 2nd paragraph) is very difficult to follow. The Documents reports the resultant BMDs in Tables 5.29 and 5.30, but fails to indicate if they are for a specific subgroup, or for the overall population. If for the latter, how the regression model is "collapsed" to represent the overall population? These two approaches are not clearly distinguished with respect to their strengths and limitations.	Both cases (a) and (b) are fully accommodated by the usual BMD procedure based on a single linear model; it is when impairment is defined based on lower limit of normal that the demographic specificity interferes, which is why the so-called "empirical" BMD procedure was applied. As for interactions representing different responses by sex or other risk factors, while perhaps medically relevant, these are not matters that are recognized in regulation. NIOSH RELs are one-size-fits-all in terms of sex, age, etc., which is why the criteria document risk assessment ignores these distinctions (except for smoking, a strong competing cause which was explored). An example has been inserted. See also response to comment 5086.
5139	Section 5.4.1.3: It is difficult to follow. On line 13, "Adding the predicted high-risk cases for different levels of exposure metric to the empirical BMD table (table 5.30?) provides composite BMDs including the high-risk contribution (Table 5.34)." Consider using an example to illustrate.	See response to comment 5138.
5140	Section 5.4.2: Although not clearly stated, it appears to be Poisson model 4 (Table 5.2.4) that was used for BMD calculation. Model 4 involves gender, age, ever-smoking and packyr of smoking. The results of BMD in Table 5.3.5 are based on two models, with or	Models are now specified with the tables.

	without smoking status. Exactly which Poisson models were used? Further, justify why other factors were considered in the model but ignored in BMD computation.	
5141	Section 5.5: To summarize the risk estimation that has been conducted under combinations of many options, a summary table of all risk estimates, like table 5.37, would be very useful for comparison purposes. Section 5.5 for sensitivity analysis would be a good place. This will tie with the introduction, and let the reader to be better informed of why certain model, subgroup, exposure metric, risk estimation approach were chosen. All tables should be checked to ensure complete information provided, including the title and footnotes.	Models are now specified with the tables.
5142	Yes, I support the use of a10-year duration of exposure to supplement a 45-year exposure. Given the turn-over rate of workers in this industry and the possibility of a higher-risk population susceptible to earlier onset of lung obstruction, short-term risk should be considered.	See response to comment 5092.
5143	I have no comments.	No response required
5144	There are occasions of typographic error and inaccuracy in this Document which can be easily fixed through a careful proof-reading. - Table 1 in Chapter 2 for example identifies exposure assessment studies by labeling company in letters and using references such as NIOSH (2008). Yet in the test (section 2.5), these studies are identified by the state where the company is located as well as by the references such as NIOSH (2008a, b) etc. This leads to unnecessary confusion. a) Similar errors are seen in Table 3.1: NIOSH [2003] should be [2003a], NIOSH[2008] should be [2008b]. b) - P55, line 11-12: "Mean diacetyl air concentrations in other plant areas were less than 0.15 ppm". In Table 1, it is ">0.15 ppm" c) - P55, line 25: "NIOSH conducted seven follow-up medical and environmental surveys at the index microwave popcorn plant …". So together with the first survey the total number survey in the index plant is eight. However, in Table 5.1, the number of surveys is 9. d) - P65, line 23: "… 12 were production workers." But Table 3.1. (pp48) (NIOSH 2008) states 14 workers.	(a) Table 3.1 has been updated to address the errors noted. (b) The reviewer noted disagreement between the text in Chapter 3 on page 55, lines 11–12 and a value displayed in Table 1 of Chapter 2. The text in Chapter 3 is correct: "Mean diacetyl air concentrations in other plant areas were less than 0.15 ppm." Table 1 was changed to reflect that the diacetyl values for "Other areas" were <0.15 ppm. (c) The reviewer also noted that the text in Chapter 3, page 55, line 25: "NIOSH conducted seven follow-up medical and environmental surveys at the index microwave popcorn plant …" contradicts Table 5.1 (Chapter 5) where the number of surveys is given as nine. A footnote to Table 5.1 clarified that there were a total of eight medical surveys and nine environmental surveys. We have added text to page 55 to echo this clarification. (d) This is not a contradiction. Testing of 12 current production workers is discussed in one location and 14 workers who ever worked in production in the other. (e) The text is correct as stated (Appendix

	e) - P117, line 13. Should be Appendix 4. f) - P212-213, "Table 5.36" should be "Table 5.38" g) - Section 6.1: references to Table 6.5 are incorrect. It should be Table 6.4. h) - P123: Table 5.5. in the last sentence should be Table 5.6 i) - Page 137, line 14: should be "45-year".	3). (f, g, and h) These typos were corrected. (i) "10-year" is correct as stated.
5145	NIOSH selected only plants K, L, and G for dose-response modeling, excluding other plants. It is unclear if and what inclusion-exclusion criteria NIOSH had. NIOSH then combined plants K and L to form a pooled analysis. The arguments (a-d) above for not pooling plant G do not hold well. Appropriate exposure metrics and exposure history are designed to capture the specific conditions of plant G concerning arguments a, b, and d; using exposure history and approximate disease onset time in a Poisson regression may capture that of d). The exposure coefficient from the regression model being lower for plant G is likely the result of higher historical exposure there, which itself can hardly be a defensible reason for exclusion of plant G. Decision for inclusion should be based on study design, not based on results of statistical analysis. It is otherwise an analogy to deciding the exclusion of a study from a meta-analysis because of being non-significant. Pooled appropriately, combined data analysis would increase the statistical power and reduce chance variation. With pooled data plant-related clustering effect can be accounted for in the model to retain the statistical capability of quantifying the differences between the plants which are otherwise difficult to characterize. One common approach is to introduce "random effects" for the plants. To this end, NIOSH needs to carefully develop	See response to comment 5097.

	and review the selection criteria and determine if the other three popcorn plants meet the inclusion criteria as well. Having lower exposure level and small number of cases, for example, are not among reasons for exclusion when these plants are in fact integrative parts of the NIOSH HHE studies.	
5146	The case definitions in the Document are informative. As pointed out in the Document 57% of the cases were not tested for reversibility on lung obstruction, but only one case among those cases that were tested was reversible, although the test using bronchodilator medication has a residual deficit. Therefore, over-diagnose (over-classification) of BO under the two case definitions is likely, but with the rate unknown, and the result would be an overestimation of risk. With still another case definition of **FEV1/FVC<LLN,** we may gain empirical evidence on the magnitude of over-classification of BO compared with the case definition **FEV1<LLN**. Together, the two less restrictive definitions would indicate upper bounds on the overestimated risk using **FEV1<LLN AND FEV1/FVC<LLN.** Additionally, it permits quantification of uncertainty. Finally if the possibility of reversible lung obstruction effects is of concern, direct use of lung function decline is well justified. In this case, it might be worthwhile to consider adopting the benchmark dose method for the continuous outcomes of FEV1 or FEV1/FVC directly (without dichotomizing them).	See response to comment 5015.
5147	Excluding the asymptomatic subjects is likely to under-diagnose lung impairment (e.g. BO), hence underestimate the risk associated with exposure to diacetyl. In contrast, including the asymptomatic subjects may over-diagnose pulmonary impairment, leading to overestimation of risk. Therefore, conducting dose-response assessment by including and excluding asymptomatic subjects respectively would show different values of the risk estimates that inform the uncertainty and inaccuracy of case diagnosis and	See response to comment 5016.

	definitions. To avoid bias, analyses of including and excluding asymptomatic subjects should both be conducted.	
5148	This is indeed a concern in modeling especially with the regression models. Given that in plant K and L there is only one environmental survey, AvgExp is perfectly proportional to duration. A perfect co-linearity would be the case if both cum(DA) and duration were used in the regression models for FEV1 and FEV1/FVC. In plant G, the multiple and changing values of DA from the multiple surveys attenuated the collinearity. The cases of early onset of a potentially higher risk subpopulation also artificially support the duration effects. Still the artifact of collinearity is implied in the models: after the Duration, cum(DA) is mostly non-significant (Tables 5-11 ~ 5.13). The significance of cum(DA) in Table 5-15 is unclear; the Document does not explain the asterisk for cum(DA), as in several other occasions. The Document neither explains adequately the issue of collinearity of cum(DA) and duration and the subsequent negative impact on duration, nor articulates the purpose of the various statistical modeling that were NOT used to derive risk estimates at the end.	See response to comment 5017.
5149	While average exposure matters, it dilutes the difference between a longer period and a shorter period of exposure at the same daily average level. The Document's use of cumulative exposure is appropriate. The additional use of average exposure in the absence of peak exposure may help us to better understand the uncertainty involved in various exposure metrics in general. We note that there were clusters of cases of early onset. These cases may be responsible for the diminishing difference between average exposure and cumulative exposure in the present NIOSH studies. This attenuated difference may be one possibility that average exposure (in conjunction with duration) appears to have contributed to the model(s). Similarity of average exposure to cumulative exposure is also attributable to the lack of daily exposure measurements over a long period.	We agree with these comments but would point out that average and cumulative would be similar measures only if durations were quite uniform.

5150	Assuming there is indeed a higher-risk subpopulation susceptible to earlier onset, explicitly differentiating such a sub-population from the overall population is highly useful. The conventional statistical approach uses a mixture distribution that combines the individual distribution for each sub-population into an overall distribution for the pooled population. The idea of incorporating the term HRX to depict a declining risk over time is a good one. However, this term is a part of the incidence rate parameter of the Poisson model and is applicable to the entire population. As a result, although the intention of the Document is to use HRX to describe a highly susceptible subpopulation that has a higher short term risk, HRX is not specific to the subpopulation. In essence HRX is an interaction term of $Cum(DA)^2$ with a nonlinear transformation of duration (i.e., $exp(-0.693Duration)/Duration^2$) in the incidence rate for the entire population. It amplifies the incidence rate and dominates the exposure ($Cum(DA)+Cum(DA)^2$) when Duration is short; it diminishes in value and let the exposure Cum(DA) nominates the incidence rate when Duration is greater. The Document should cite relevant references for HRX. The Document correctly points out the larger number of cases in excess of what a Poisson distribution anticipates. It might be helpful for NIOSH to investigate if group 2 workers in the index plant were responsible for most cases of earlier onset. The mixture modeling approach to a pooled population would remain interesting and useful. It is difficult to judge if the person-year exposure for earlier onset cases is large enough to be “robust”, and a post-hoc assessment is required and also seems feasible.	See response to comment 5019.

5151	The Document presents a cohesive approach to the analysis of the animal data including the estimation of benchmark doses. The exposure-response modeling of the animal data is adequate and acceptable in general. There are some concerns that invite further discussion and perhaps some alternative approaches. Data of 6-week and 12 week exposure experiments were compared and then combined because the incidences of lesions are not statistically different. Statistical significance can be determined by the number of animals involved and the two experiments are of very small size with 4 groups and 5 animals per group. So small sample size could have contributed to the non-significance. If feasible, justification should be made from a toxicokinetic/toxicological standpoint or even an experimental standpoint. It is unclear from the Document if cumulative exposure was considered when analyzing the combined data. Assuming a cumulative exposure was not used, it is a viable alternative to consider the cumulative exposure. Such an alternative is attractive because 1) it would be consistent with exposure metric in the occupational epidemiological study and 2) it captures exposure duration for the two experiments. The dose-response data of nasal inflammation and peribronchial inflammation do not support the use of benchmark response (BMR) level at 0.001. There is little dose-response data in this range. As a result all exposure-response models considered in the Document are highly uncertain in the exposure-response shape up to the lowest exposure level. To what extend this uncertainty is responsible to the engulfing difference between the BMDs for penetration and tissue concentrations is unknown. But instead of using BMR=0.001, NIOSH should consider BMR=0.05 or 0.01 as a point of departure (POD) to help quantify the uncertainty. If BMR=0.001 is to be used, a brief discussion on the uncertainty at this BMR level would be useful, and so is a comparison with linear extrapolation from a POD at higher BMR levels. On pages 191-192, the Document provides a detailed discussion of variation/uncertainty of BMD and BMDL in association with various factors, including model choice, scrubbing factor, VE, breathing rate, etc. The range of BMD and BMDL estimates	The BMD analysis presented in Chapter 6 of the draft document has been updated on the basis of new information and replaced by a categorical analysis, as described in the revised Chapter 6. The updated analysis is based on a response rate of 10% as opposed to 1 in 1,000; therefore, the portion of the comment referring to a BMR of 0.001 is moot. The categorical regression model in the updated version of Chapter 6 does include a term for the duration of exposure; therefore, a statistical adjustment was made for the differing durations of the exposures in the animal studies. This adjustment was based on the observed differences in the toxicological responses to exposures of 6, 12, or 13 weeks, and is more consistent with the data than a simple assumption of a response proportional to the cumulative exposure would be.

	corresponding to these possible options informs well the scope of uncertainty and variability at this end of risk assessment.	

Follow-up review of new content (revised Chapter 6 and new section of Chapter 8), December 2013

Tracking Number	Reviewer's comment	Response
7001	No. The use of single-point "uncertainty" factors fails to adequately represent quantitative information about either human variability or the uncertainty in interspecies projection of internal doses and effects. It also fails to produce quantitative estimates of human risks of adverse effects for the existing distribution of exposures or to the exposure distribution likely to result after the implementation of alternative OSHA standards for worker protection. This failure prevents OSHA from adequately addressing its needs for quantitative benefits analysis under OMB guidelines for evaluation of alternative standards and under the old benzene decision mandate for OSHA to demonstrate "significant risk" of material impairment under current standards that would be reduced with a candidate alternative standard.	NIOSH disagrees. The application of uncertainty factors is explicitly intended to allow for both human variability and uncertainty regarding interspecies differences. The use of uncertainty factors represents a practical and widely-accepted method for establishing occupational exposure limits from toxicological data. In regard to the assertion that the use of uncertainty factors would prevent OSHA from conducting a quantitative benefits analysis, NIOSH notes that such an analysis could readily be based on the analysis of human data presented in Chapter 5.
7001a	Following review of the analysis of the human data in Chapter 5, I believe that the extensive human observations of lung function changes in workers exposed to diacetyl are a much better starting point for analysis of human risks and selection of desirable human exposure standards. The human observations appear to document a serious amount of variability among people in their sensitivity to lung function impairment from diacetyl, with a minority experiencing appreciable changes with short exposure times. Although the lung changes are different, this is reminiscent of the berrylium case, where proteciton of the more susceptible workers requires very severe restrictions on permitted levels in the workplace. The current regression based analysis in Chapter 5 can and should be improved. In particular there should be imputation of the mean values for the substantial numbers of non-detect observations of diactyl levels in different plants by fitting lognormal distributions to the portion of the data above the detection limits (I can amplify on the exact methodology for doing this if I am given the individual data set for one of the plans, preferably plant G, which has the largest amount of	NIOSH agrees that the analysis of human toxicity presented in Chapter 5 is a preferable basis for developing occupational exposure guidelines than the analysis of animal data presented in Chapter 6. NIOSH has added additional language to Chapter 6 to emphasize that the proposed NIOSH REL for diacetyl is in fact based on the human data, and that the analysis of the animal data is intended only as supporting evidence for the REL.

	data). However even as it stands, the Chapter 5 analysis represents an extensive and creative application of risk assessment techniques to the data that I believe is greatly preferable to the use of projections of animal data for this purpose. Because of likely significant differences in the deposition sites for the rodents vs people (because of nmore extensive scrubbing in the upper airways) it is unlikely that the rodents can provide reasonable proxies for the deep lung effects of an agent with the solubility of diacetyl. The human data do have complications in the form of co-exposures to other lung toxicants, but I think these are likely to be less serious difficulites than those inherent in the projections from rodent data.	
7002	The current document's description of the endpoints, the severity scale, and the underlying dose response information modeled are all inadequate to develop a reasoned response to this question. The human effect of concern occurs deep in the lung at the level of the bronchioles, but the distribution of the sites of absorption in the rat are expected to be higher in the respiratory tree because of more efficient scrubbing and possibly metabolism related removal in the upper airways in rats compared to humans.	NIOSH agrees that there is considerable uncertainty in extrapolating from diacetyl toxicity in rodents to diacetyl toxicity in humans. NIOSH has attempted to address this uncertainty through the use of a physiologically-based pharmacokinetic (PBPK) model to address species differences in the sites of toxicity and by the application of uncertainty factors to address other sources of uncertainty, such as interindividual variability and the effects of chronic exposure. It should be noted that the proposed REL for diacetyl is based on an analysis of diacetyl toxicity in occupationally-exposed workers, and not on extrapolation of diacetyl toxicity from rodents to humans.

7003	The model described is useful as far as it goes. However there is no representation in the current model of the effects of variation among humans either in anatomical details or in metabolism rates. Even the metabolism rates used in the model appear to be based on rat data, and this could introduce a bias because many metabolism rates tend to scale with body weight to the (-1/4) power (we bigger animals tend to metabolize less per unit body weight by about this factor). In this case a lower metabolism rate per unit tissue mass would be expected to lead to still greater penetration of diacetyl to the deep lung region where the damage occurs in people than would be expected from the current version of the model. That this issue is not discussed and reflected in some distribution-based adjustment to represent the associated uncertainty in the dosimetry is unfortunate, and should be corrected before the document is finalized.	The reviewer's assumption that the diacetyl metabolism rates for human tissues in the Gloede et al. [2011] PBPK models are based on rat data is not correct. A close reading of the model description indicates that diacetyl metabolism in the respiratory epithelium was in fact scaled from rat to human on the basis of experimental data. The Gloede et al. PBPK model does clearly predict deeper penetration of diacetyl in the human respiratory tract than in rodents; therefore, NIOSH believes that the model adequately accounts for the impact of this interspecies difference.
7004	This is clearly indicated.	A bronchiolar human equivalent concentration (HEC) has not been added, because the rodent data for diacetyl do not provide a bronchiolar point of departure usable for risk assessment.
7005	As indicated in my response to question 3 above, the model needs to be improved by distributional assessment to represent not only the variability among people in metabolism rates but also the uncertainty in the use of animal-based measurements of metabolism to estimate human metabolism without any apparent adjustment. Absent any other data I would adjust the central estimate of human metabolism rates with the body weight^(-1/4) factor alluded to above, but also show distributional implications of likely human variability based either on direct observations of the variability among humans in likely metabolism rates, or variability that could be inferred from earlier data available in the literature. One starting point for such an analysis could be an older paper that I published on likely variability in dosimetry and response to airborne particulates: Hattis, D., Russ, A., Goble, R., Banati, P., and Chu, M. "Human Interindividual Variability in Susceptibility to Airborne Particles," Risk Analysis, Vol. 21(4), pp. 585-599 (2001).	As noted above, the reviewer's assumption that the diacetyl metabolism rates for human tissues in the Gloede et al. [2011] PBPK model are based on rat data is not correct; therefore, no allometric adjustment of metabolism is required. NIOSH has adjusted for human variability through the use of an uncertainty factor, described in section 6.2.2.6.

7006	I have not been provided with this Appendix in the materials sent to me for review. Also, I would note that mouse breathing rates are importantly related to the concentration of irritating chemicals they are exposed to. To the extent that projections are being made from the mouse data (as opposed to the data from rats, which do not appear to reduce their ventilation in response to irritant concentrations) then this must be taken into account in making inferences of the tissue concentrations that gave rise to the observed effects.	NIOSH agrees that mouse breathing rates are related to the concentration of irritating chemicals they are exposed to and has adjusted for this effect by including a quadratic term in the BMD model for mice.
7007	I have now been provided with Chapter 5 for review and, as indicated in my response to question 1, I believe the human data represent a far superior basis for risk assesement than the rodent data. In any event, I think the days of selecting arbitrary single-point "uncertainty factors" instead of distributional data or distributional estimates I think should be long past. Uncertainty distributions for the current analysis should be derived by (1) whatever comparisons are possible between the animal and human data on dosimetry and effects, such as metabolism rates, and (2) data or estimates of the human variability in both dosimetry and likely differences in susceptibility per unit of dose delivered to relevant anatomical sites. The subchronic/chronic factor in this case is even more subject to uncertainty than usual because my impression of the human data is that the effect builds up over a period of years wheras evidently the animal data did not detect a difference in apparent response between exposures of the order of 6 and 13 weeks. Overall, I am not optimistic that the animal data add much to what might be more directly inferred from the observations of chronically exposed workers.	NIOSH agrees that human data provide a superior basis for diacetyl risk assessment than the toxicological data. The proposed NIOSH REL for diacetyl is in fact based on the human data, and the toxicologically-based risk assessment for diacetyl is intended only as supporting evidence for the REL. Language has been added to Chapter 6 to clarify this point.

7008	I have also received from NIOSH a recent article on the association between declines in lung function in exposed workers in relation to diacetyl exposure: Am J Ind Med. 2014 Feb;57(2):129-37. doi: 10.1002/ajim.22282. Epub 2013 Nov 22. Work-related spirometric restriction in flavoring manufacturing workers. Kreiss K. Author information Abstract BACKGROUND: Flavoring-exposed workers are at risk for occupational lung disease. METHODS: We examined serial spirometries from corporate medical surveillance of flavoring production workers to assess abnormality compared to the U.S. population; mean decline in forced expiratory volume in one second (FEV1) and forced vital capacity (FVC); and excessive declines in FEV1. RESULTS: Of 106 workers, 30 had spirometric restriction, 3 had obstruction, 1 had both, and 13 (of 70, 19%) had excessive declines in FEV1. The adjusted prevalence of restriction was 3.7 times expected. Employees with higher potential for flavorings exposure had 3.0 times and 2.4 times greater average annual declines in FEV1 and FVC respectively, and had 5.8 times higher odds of having excessive FEV1 declines than employees with lower potential for exposure. CONCLUSION: Exposure-related spirometric abnormalities consistent with a restrictive process evolved during employment, suggesting that exposures in flavoring production are associated with a range of pathophysiology. Am. J. Ind. Med. 57:129-137, 2014. Published 2013. This article is a U.S. Government work and is in the public domain in the USA. Published 2013. This article is a U.S. Government work and is in the public domain in the USA. KEYWORDS:	NIOSH performed the cited study.

	diacetyl, excessive decline, flavorings, hydrogen sulfide, spirometric restriction, spirometry	
7009	Unfortunately, this paper does not provide data in a form that is immediately helpful for risk analysis. Re-analysis of distributions provided in the data underlying this work may be helpful, however.	Data on the decline in lung function in diacetyl-exposed workers is in fact the basis of the NIOSH risk assessment for diacetyl that forms the basis of the NIOSH REL.
7010	Yes. Absent a probabilistic treatment, the projections cannot address OSHA's needs to quantify how many workers are likely to be at significant risk of material impairment, and the uncertainty in how much health benefit could be realized by reducing the current exposure distribution by various amounts.	NIOSH believes that estimates of the numbers of workers at risk for material impairment due to diacetyl toxicity should be based on the modeling of human data that is presented in Chapter 5 of the document.
7011	Absent other information, this seems a reasonable inference. However I should note I have big problems with the current presentation of the modeling methods and results. I made the following notes when going through the current chapter 6 draft:	See responses to specific comments below.
7012	p. 13, bottom through p. 15 top—it is crucial for the document to clarify and justify for the reader the implications of the multiple logistic equation for the assumption about the form of the dose response relationship. No justification—mechanistic or otherwise—is now provided. This is now presented as a purely statistical exercise. Much more is needed to convince toxicologists or those with mechanistic orientation that the analysis is reasonable.	Support for the ordinal logistic model has been added including descriptive text and supporting citations of the literature on its merits for application to these data.
7013	P. 15, bottom. The reference to a "negative quadratic effect" is particularly opaque and needs to be clearly explained and illustrated with a specific equation, fitted parameter values, and confidence limits.	These items have been added to Chapter 6.

7014	p. 16. Three random effects parameters and 53 fixed effect parameters is surely excessive. There needs to be a careful evaluation and explanation of how the data leads to each one, if it does, and what they all mean. There is an old saying, "With four parameters I can fit an elephant; with five I can make it wave." That use of 56 parameters seems far beyond the pale of reasonability.	We acknowledge the concern expressed about the number of parameters. We recognize that the description of the model and its parameters was incomplete as it inadvertently omitted a few substantial details. These details have been added, and the text has been revised to improve clarity. Although the specific context for the reviewer's example wasn't specified, models that contain more than five parameters are not uncommon, e.g., PBPK models. The issue is whether the information contained in the data is sufficient to identify all of the parameters. Two lines of evidence relevant to this issue were subsequently developed, (a) based on the Hessian matrix of the algorithm for optimization and (b) profile-likelihood evaluations in a neighborhood of the solution. The evidence developed from both lines of evidence support a conclusion that our model was not over-parameterized although a strong assumption about the effect of exposure duration among females was necessary to avoid an over-parameterized model of the females. Chapter 6 has been revised accordingly.
7015	p. 17, 1st paragraph. This reads as gibberish. Basic explication of concepts and toxic mechanism-based implications are essential to convince anyone with a toxicological background that the statisticians have not just completely gone off the deep end here. The entire multi-step process of the modeling needs to be described with equation forms, data, empirical results of model fitting and choices for the next step in the modeling provided in a transparent text with tables, figures etc. The current jargon-filled overview paragraph is just not acceptable.	The text was intended to be a technically accurate and concise description accessible to statisticians. However, additional description and citations have been added to Chapter 6 to support the analysis and improve accessibility although it is nevertheless technically dense and admittedly a challenge to fully comprehend. Additional results summarizing the multistep process of the modeling will be made available upon request.
7016	p. 18, 1st paragraph. The rat endpoints which could be 7 modeled adequately according to the criteria listed in section 6.2.2.1 (a score test for separate 8 slopes and a likelihood ratio test for an unrestricted multinomial distribution) are shown in Table 6.6. Mouse endpoints which could be modeled adequately by the criteria described in section 6.2.2.7 are shown in Tables 6.7 and 6.8.	7016 has been combined with 7017. See response to 7017.

7017	The criteria for “adequate” modeling are not clearly laid out in the sections referred to. And in any case it is not clear that inadequate modeling should be a cause for throwing out the data/endpoints. Perhaps it is the overly complex modeling approach that needs to be modified instead. In any case, transparency requires not just a qualitative listing of endpoint responses in Tables 6.1 and 6.2, but a reasonably accessible presentation of the underlying data that is to be modeled, with simple statistical evaluations such as significance tests vs control observations.	The criteria for deeming a model as “adequate” are fully specified; hence, NIOSH disagrees about these being unclear. However, NIOSH agrees that models deemed inadequate are candidates for revision but it is likely that such revisions would require a more complex model. NIOSH obtained these data from the original investigators who should have control over its dissemination. The suggestion to perform multiple pairwise tests vs. control observations would be an inefficient use of the data and yield a large number of tests that invites complications for their interpretation. Tests of a null dose-response within the ordinal logistic model were performed as they use all of the information in the ordinal response data and would enjoy superior power for a wide range of alternatives.
7018	p. 18 “The BMC and BMCL estimates were extrapolated to HECs as described in sections 6.2.2.2 – 6.2.2.4, and the HECs were converted to candidate REL values by the application of UFs as described in section 6.2.2.5. The BMC/BMCL values for rats, and their corresponding HEC and candidate REL values are shown in Table 6.6. The BMC/BMCL values for mice, and their corresponding HEC and candidate REL values are shown in Tables 6.7 and 6.8; the BMCL values in Table 6.7 have not been adjusted for overdispersion, while the BMCL values in Table 6.8 have been adjusted for overdispersion. The criterion given by Bock [1975] supported making no minimax adjustments of these estimates”	See response to 7019.

7019	It is unclear what the authors mean by "overdispersion" (presumably some greater observed variance than expected, but in which parameters, and how measured?) how this is manifested in the data and what adjustments exactly were made in the light of this.	The presumption is correct; residual deviations were constructed, and 13 of 359 in excess of three standard errors were observed suggesting that the expected variance under the model is too small. If its cause had been isolated to a subset of the parameters then an extension of the model should — at least in theory — have been able to account for it. However, the only identified cause of the larger than expected residual deviations was extraneous variability and a straightforward extension of the model to account for it was to rescale the model-based variance matrix by multiplying it by a dispersion factor. Its impact on the results was to increase the width of the confidence intervals to avoid making interpretations solely on the basis of model-based confidence limits that appear to be vulnerable to questionably optimistic precisions.
7020	pp. 19-20, Table 6.6. First, it is absurd to present P values to four significant figures, as if the third and fourth figures had some meaning relevant to interpretation of the results. Second, The table fails to provide the basic information on what the quantitative results are in terms of dose response slopes and confidence limits for different types of responses at specific levels of severity, although it is helpful to at least provide central estimates and lower confidence limits on the benchmark concentrations.	*P* values have been revised. Tables 6.6 and 6.7 have been revised to clarify the associated response severity level and benchmark response level for the associated BMCs. Because the dose response slopes are inversely proportional to the BMCs they are redundant and have not been added.
7021	p. 24, 1st par—"The criterion given by Bock [1975] supported making no minimax adjustments of these estimates." The document is completely unclear about what "minimax adjustments" are, what the Bock [1975] criteria are, and how the data failed to meet the needed criteria. The authors just cannot rely on the reader to remember or retrieve the Bock paper to understand what is meant here, let alone interpret its importance for understanding the results.	The adjustments described by Bock [1975] had no effect on our estimates and have been deleted from our methodology because they were not relevant. However, a footnote at the end of the Discussion mentions it because those who are aware of the issue would appreciate that it was considered.
7022	p. 26—"Speculatively, reduced respiration at high exposure concentrations may contribute to the attenuation of response noted in the high exposure groups, relative to the modeled response. A strategy was therefore employed of modifying the model structure by including a quadratic dose term in modeling the mouse data, which	No response required.

	allowed sufficient model flexibility to accommodate the attenuation of response seen in the high-dose mouse data."	
7023	So rather than use known information about reduced respiration rates at high exposures to irritating gases in the mouse, and modifying the pharmacokinetic model accordingly, the authors chose to fudge their dose response model by introducing a quadratic term so the data will fit it. Bad. Not acceptable.	It is clear from the data presented by Morgan et al. [2008] that the reduction in respiration in mice exposed to high concentrations of diacetyl declined with increasing duration of exposure. Therefore it is not clear what percent reduction in respiration would correlate with the toxicity observed at the end of the 13-week NTP study. Would this be based on the average reduction in respiration observed over the 13 weeks, or the reduction observed at the end of the 13 weeks, or some other value? Therefore NIOSH believes it is preferable to treat the magnitude of this reduction as an unknown value to be estimated from the toxicity data, rather than a known quantity that could be used as a dose adjustment factor. It should be noted that the magnitude of the respiratory depression estimated from the coefficient for the negative quadratic term in the model is intermediate between the degree of respiratory depression observed by Morgan et al. [2008] at 6 and 12 weeks, and is therefore consistent with the observations of Morgan et al.
7024	p. 33—The section ends without a quantitative comparison of the animal based projections with the human findings. Some quantitative comparison of this sort, with attention to issues of uncertainty and variability in quantitative form, is essential. This is the only way to correct for possible distortions that arise from the use of single-point factors for interspecies and intraspecies projections in this analysis, and convert the results from "benchmark dose" form to quantitative projections of risk at various levels of severity. This latter is the information required for OSHA to do the analysis it needs to support assessments of "significant risk" and the benefits of incrementally more stringent standards in the context of exposures of various durations including a working lifetime.	This comparison has been added to Chapter 7.

7025	The risk estimation based upon the additional data is an appropriate reflection of the current understanding of the scientific information on diacetyl. With respect to 2,3-pentanedione (PD) and the conclusion regarding greater potency than diacetyl, this finding has been strengthened by the Zaccone et al (2013) publication where it was clearly demonstrated that PD exerted a greater effect on tracheal airway reactivity of rats inhaling either diacetyl or PD. Zaccone, E.J., Thompson, J.A., Ponnoth, D.S., Cunpston, A.M., Goldsmith, W.T., Jackson, M.C., Kashon, M.L., Frazer, D.G., Hubbs, A.F., Shimko, M.J. and Fedan, J.S. 2013. Popcorn flavoring effects on reactivity of rat airways in vivo and in vitro. J Toxicol Environ Health A 76: 669-689.	NIOSH concurs, and a reference to this study has been added to Chapter 4.
7026	Diacetyl or PD induced pulmonary toxicity in humans is associated with a chronic inhalation exposure to these agents. Although physiologically the rat serves as a more relevant model and a PBPK/CFD model is available for the rat, simulation of human exposure is more relevant from data generated in a chronic exposure model. The more appropriate parameter would be the female chronic bronchial inflammatory response in the mouse for derivation of the REL. Another factor to consider is the question of body mass index (BMI). Presumably the rodent is healthy and young i.e., aged rodents were not exposed to diacetyl or PD while in the human data the workers were presumably overweight and may have respiratory difficulties resulting in less breaths and volume, etc. In addition, smoking may also constitute a confounding factor present in humans and not the rodent model. Hence the inclusion of “light exercise” may not be appropriate as an assumption.	In the absence of job-specific data, NIOSH conventionally assumes that workers will have a respiratory rate equivalent to the International Commission on Radiological Protection “Reference Man” estimate for light activity, which is a minute volume of 20 liters per minute. This estimate is derived from a composite of light exercise and some sedentary activity, which was incorporated into the PBPK model for occupational exposure to diacetyl as described in section 6.2.2.3. NIOSH believes that it is generally appropriate to assume that the respiratory rate during occupational exposures will be greater than that for sedentary activity.
7027	As indicated above the disadvantages include the use of light exercise as not appropriate as humans varied in BMI which affects breathing rates and volumes. The factor of smoking was not apparently considered amongst the workers. In addition as opposed to the rodent genetic makeup where the animal is healthy, the presence of asthma or other diseases in humans leads to enhanced vulnerability which is not reflective in the rodent-to-human respiratory tissue concentration ratio. The advantage of this model is that data are provided to correlate a tissue concentration with an effect. However,	NIOSH agrees that the possibility of increased susceptibility to toxicity due to individual factors is of concern, as is the potential for human exposures to continue over a working lifetime. NIOSH has attempted to address these concerns by incorporating uncertainty factors for interindividual variability and chronic exposure as described in section 6.2.2.6.

	it needs to borne in mind that in humans it is a chronic exposure over years, whereas the rodent data is clearly NOT over a lifetime working exposure. The ability to induce a similar pulmonary response in an acute model is thus an advantage. The important factor may be the actual concentration of diacetyl or PD at the tissue site rather than the inhaled quantity of gas.	
7028	Based upon the fact that the human disease outcome is bronchiolitis obliterans it would be appropriate to add HEC for the bronchiolar regions. Data generated may clearly demonstrate that actual tissue concentrations and not respiratory volume or rate is the key parameter for tissue damage; this is a crucial element.	Although bronchiolitis obliterans is of concern, the rodent data do not provide a bronchiolar endpoint suitable for extrapolation to human bronchiolar disease. Diacetyl toxicity in the rodent is seen primarily in the upper respiratory tract. However, the currently available PBPK model for diacetyl [Gloede et al. 2011] predicts that diacetyl will penetrate more deeply in the human respiratory tract than in the rodent respiratory tract. Therefore, NIOSH considers it appropriate to use the PBPK model to extrapolate from the concentrations of diacetyl observed to produce toxicity in the respiratory tracts of rodents to estimated exposure concentrations in the respiratory tracts of occupationally exposed workers to estimate the toxicity of diacetyl to humans. The factors described in Table 6.4 for rodent-to-human extrapolation of airway tissue concentrations of diacetyl include estimates of the greater toxicity expected in the lower respiratory tracts of humans in comparison to rats and mice.
7029	The other factors that may be useful to employ are an “aged” rodent model or an “obese’ rodent model in an attempt to simulate the worker over a lifetime exposure. This would seem to be a more appropriate model than the light exercise	Toxicological data are not currently available for aged or obese rodent models; however, NIOSH believes that the possible effects of age and/or obesity on diacetyl toxicity would be included in the uncertainty factor for interindividual variability described in section 6.2.2.6.
7030	In my opinion additional analysis basing dose adjustment factors on measured mouse minute volume would not markedly enhance the present risk assessment.	NIOSH concurs.
7031	The selection and magnitude of the various uncertainty factors are appropriate based upon EPA criteria and adequately described. Other factors that have been alluded to include the BMI which significantly affects respiratory rate and volume. Genetic makeup such as diseases including asthma or smoking would affect intraspecies variability and	NIOSH agrees that these are legitimate concerns, but believes that the possible effects of disease and/or obesity are addressed by the uncertainty factor for interindividual variability described in section 6.2.2.6.

	account for susceptibility to impaired pulmonary function in exposed microwave popcorn workers.	
7032	In my view the addition of a quantitative model-derived risk assessment of rodent respiratory tract toxicity as a complement to the primary exposure response assessment would not add significant value. At present it has not been established that the quantity inhaled rather than the concentration at the tissue site is the critical component.	The Gloede et al. [2011] PBPK model used for extrapolation from rodents to humans does include estimates of the relative tissue concentrations of diacetyl in rats and humans exposed by inhalation. These estimates have been used in extrapolating from rodents to humans, so that diacetyl concentrations in the target tissues have been incorporated into the risk assessment to the extent possible given current data.
7033	As indicated previously in the Zaccone et al (2013) publication it appears that PD exerts greater effects on bronchial airway reactivity that diacetyl. These observations indicate that PD is in fact more potent than diacetyl and support the NIOSH conclusion that PD needs to be treated as equi-or greater in potency than diacetyl.	NIOSH concurs.
7034	Thank you for the opportunity to review this Subsection of Chapter 8. Firstly, let me say that the document is a well constructed, carefully crafted communication and the author(s) are to be commended for their competency and painstaking attention to detail. Before I address the specific questions asked I have the following suggestions / comments which may be helpful in preparing the final document. General Suggestions and Comments:	No response required
7035	Section 8.3.7.1 line 7. States “Employers should establish a comprehensive safety and health ***training*** program …etc” . Then, on lines 10 and 11 the document mentions “airborne monitoring” and “medical surveillance” ***as though they were part of*** the training program. I suggest the anomaly be clarified by removing the word “***training***” from line 7.	The word “training” was removed.
7036	Section 8.3.7.3 line 16. This CFR is missing from the reference list.	This reference was added to the reference section.
7037	Section 8.3.7.3 line 21 The Cal/OSHA 2013 is missing from the reference list.	This reference was added to the reference section.
7038	Section 8.3.7.3. Somewhere in lines 24 to 28 should we actually say what the (proposed) NISOH REL is, for the benefit of those who don’t know, or don’t memorize these things?	The NIOSH REL and STEL were added to the text in this section.

7039	Section 8.3.7.3 lines 13 and 14. I think we need to be a bit careful how we say this. We say "…*the word 'warning' in the FEMA text conflicts with standardized GHS terminology."* This isn't exactly true is it? The word "Warning" is a Bona Fide term in the GHS. I think what we are trying to say is that the word "Warning" in the FEMA text is: ***inconsistent with the specific criteria for its use and application as a Signal Word in the GHS.***	The reviewer's revised text suggestion has been added to the revised document.
7040	Even with the above clarification, we tend to leave the reader "up in the air" and exasperated. We've told them of the problem with the FEMA text, but not what to do about it. It's not until they arrive at the next section that the solution is presented. I suggested we give them a "heads up" by ending with statement like: *NIOSH recommends removal of the word "warning" when using the FEMA text (see section 8.3.7.4 for details).*	The reviewer's revised text suggestion has been added to the revised document.
7041	Section 8.3.7.4 lines 11 and 12. The explanation of the problem with FEMA's use of the word "warning" is a little clearer here, but still not quite on the mark.	The text suggested by the reviewer in comment 7039 has been added here as well.
7042	References, line 12 Hubbs, Cumpston et al 2012. Should the 's' and 'd' of Sprague-Dawley be capitalized?	The CDC style guide indicates that proper nouns should be capitalized in references. No change was made to this reference.
7043	References, line 23 OSHA 2013 – despite the fact that the word "diacetyl" is in the url I could not find any obvious mention of this guidance document at the web address.	The OSHA webpage "Hazard communication guidance for diacetyl and food flavorings containing diacetyl" has several sections that discuss hazard communication standard guidance for diacetyl, based on the original OSHA hazard communication standard (not the revised standard). No changes were made to the document based on this comment.
7044	References, line 30 Sigma Aldrich. The url did not lead me directly to the appropriate SDS. The relevant SDS can be found at http://www.sigmaaldrich.com/catalog/product/sigma/d3634?lang=en&region=US. A note at this site suggests Sigma has discontinued the product.	The Sigma Aldrich Web address provided by the reviewer was updated in the document.
7045	I believe I only have a revised "portion" of Chapter 8, which is probably why I could not find the following references cited in the text: NTP 2011; NIOSH 2008b and NIOSH 2009b	NTP 2011 was located in Table 8.2. NIOSH 2008b and NIOSH 2009b were located in section 8.3.7.3.

7046	I have carefully checked the hazard classifications against the cited peer reviewed literature. In my opinion these classifications are entirely reasonable and accurate. Even though there are detailed and rigorous criteria for classification under the GHS, there will always be a need to exercise professional judgment in the final analysis. An example is the classification of diacetyl as a Category 2 flammable liquid. The flashpoint data for diacetyl is contradictory, ranging from 45F (7C) to 80F (27C) depending on the reference source and test methodology used. In the higher flashpoint range of 80F (27C) diacetyl would be in Category 3 not 2. The fact that NIOSH chose Category 2 reflects that professional judgment, has been used appropriately. The classifications reflect a very reasonable interpretation and understanding of the data presented.	We further reviewed the flashpoint cited by the National Fire Protection Association (NFPA) for diacetyl and contacted NFPA to determine the primary source for this information. NFPA indicated that they needed to further review their published flashpoint based on the other flashpoint data that was available. Given this information, we decided to retain the flashpoint data that was cited as part of the draft GHS classifications.
7047	I believe NIOSH has more than adequately explained why the cut-off values should be lowered with respect to diacetyl. However, I think the authors have been so focused on the justification, they forgot to articulate the actual goal. Nowhere in section 8.3.7.3 can I find a clear, unambiguous statement on the lowering of cut-off values for diacetyl. Perhaps I am misinterpreting this part of the question. As for extending the approach to 2,3-pentanedione, this is a reasonable precautionary measure. The justification is perhaps tenuous, given the difference in vapor pressures, but the knowledge that 2,3-pentandione may be heated during processing is, in my opinion, sufficient to merit the caution. I searched the literature extensively looking for additional information indicating the volatility of 2,3-pentandione justified the NIOSH approach, but there was little to be found. However, one interesting publication - *Rincon-Delgadillo M.I. et al. [2012]. Diacetyl levels and volatile profiles of commercial starter distillates and selected foods. J. Dairy Sci 95(3):1128-1139* - reports 2,3-pentanedione in the vapor headspace studied in their investigation. (See Table 3, page 1133 in the publication). The data is sparse, but it clearly demonstrates the potential for vaporization in commercial operations.	Additional language has been added to this section to clarify the NIOSH recommendation to provide classifications below the default cut-off values. Regarding the additional journal article provided by the peer reviewer [Rincon-Delgadillo et al. 2012], we appreciate the reviewer alerting us to this reference. We critically reviewed this information for relevance as further evidence for the cut-off limits provided in the GHS sub-chapter. In our judgment, we did not think that the data presented in this article could be used to further justify a lower cut-off point for GHS classification of 2,3-pentanedione.

7048	Most people will be confident with the NIOSH derived classifications, without the derivation details. If they need to backtrack into the derivations then the literature citations in Tables 8-2 and 8-3 are easy enough to find on the net. That being said, the dominant audience for this document is probably the "Technical Specialist". These professionals, by their nature, usually want as much detail as possible. If there is a significant demand, perhaps details of the derivations could be placed in a table as an addendum to the Chapter. The table would have 4 columns. Column 1 would contain the hazard classification and category. Column 2 would have the literature citation(s) used to make the classification. Column 3 would have the end point data (such as LC50, Flash Point etc) from each citation, and Column 4 would have the rationale for the GHS classification, based on the data in Column 3. Placing such an easy-to-read table as an 'addendum' would avoid cluttering the chapter with extraneous detail.	The rationale for the GHS classifications has been added to Tables 8.2 and 8.3 in Chapter 8.
7049	Thank you again for the opportunity to review this most interesting document. I trust my comments and suggestions will be helpful.	No response required
7050	Section 8.3.7.2, GHS Classifications of Diacetyl and 2,3-Pentanedione. NIOSH's GHS classifications do reflect a reasonable interpretation of the presented peer-reviewed data. However, while NIOSH notes (on page 3, lines 4 through 7, that "Appendix C of the hazard communication standard [29 CFR 1910.1200] provides several precedence rules regarding the application of pictograms and signal words as well as rules for combining or omitting hazard and precautionary statements," it would be in the public's interests for NIOSH to add another paragraph in between Tables 8-2 and 8-3 which presents the combined results, e.g., one Signal Word ("DANGER"), the appropriate pictograms (Skull and Crossbones, Corrosive, Health Hazard and Flammable Liquid, but no Exclamation Point) and a combined set of Hazard Statements and Precautionary Phrases. In other words, NIOSH needs to provide a complete set of Label Elements for the regulated community to use.	We have provided all of the pertinent information in our GHS classifications that will allow manufacturers to create labels. We do not think it is necessary to create a complete set of labels in the criteria document.

7051	Same comments (NIOSH needs to provide a complete set of Label Elements) as above apply to what should appear after Table 8-3 for 2,3-pentanedione.	We have provided all of the pertinent information in our GHS classifications that will allow manufacturers to create labels. We do not think it is necessary to create a complete set of labels in the criteria document.
7052	The information NIOSH provides regarding diacetyl and 2,3-pentanedione in which these compounds which might be present in mixtures below one or more cut-off levels might still present hazards (page 4, line 19, over to page 5, line 8) is sufficient upon which to base NIOSH's conclusions that mixtures containing these materials present below the cut-off limit for respiratory single exposure and repeated exposure toxicity, still present hazards to workers. As a result, manufacturers preparing such mixtures must provide appropriate hazard communication information in their GHS-compliant SDSs and must be label such mixtures accordingly. However, NIOSH could (and should) provide more direct guidance to such mixture manufacturers as to how such mixtures should/must be labeled.	Tables 8.2 and 8.3 were updated with footnoted information to provide additional guidance to the reader about specific cut-off values/concentration limits for specific endpoints.
7053	While the Flavor and Extract Manufacturers Association (FEMA) has suggested some precautionary statements (page 6, lines 2 through 5), the suggested language is only appropriate for MSDSs and not for GHS-compliant SDSs. For GHS-compliant SDSs, NIOSH, at the end of Section 8.3.7.3, should emphasize the importance of mixture manufacturers advising users of the hazards for specific target organ toxicity – single exposure and specific target organ toxicity – repeated exposure by providing a table – much like what is included in Tables 8-2 and 8-3 – with the appropriate signal word (DANGER), the appropriate pictogram (Health Hazard) and the appropriate hazard phrases for these foreseeable conditions of use. In order to protect users, NIOSH needs to do everything it can do to give mixture manufacturers specific guidance for these hazards for both compounds.	We feel that this information is well described in the document so an additional table describing this information again is not necessary.

7054	Almost. People familiar with GHS classification procedures can relate to the GHS endpoints shown in the leftmost column in Tables 8-2 and 8-3 and can look up the reference in under the second column if they need further information. For those people not familiar with the hazard classification process, NIOSH could add, as a new first paragraph to Section 8.3.7.2, GHS Classification of Diacetyl and 2,3-Pentanedione, a two or three sentence paragraph which describes specifically where (Appendix A) the various endpoints are discussed by Appendix A section number. Again, the suggestion is made so that the reader of this Chapter of the Criteria Document is perfectly clear on where the requirements for classifying diacetyl and 2,3-pentanedione can be found.	An additional sentence referring the reader to Appendix A (health hazard criteria) and Appendix B (physical hazard criteria) were added to section 8.3.7.2.
7055	Another suggested change would be to include, in the last bullet point of section 8.3.7.4, Labeling and Posting, the introductory phrase, "As soon as possible, but no later than June 15, 2015, [F]follow the requirements of the HCS…" NIOSH should say that, given the importance of warning users about exposures of diacetyl and 2,3-pentanedione above the REL, that development of GHS-compliant SDSs and labels should happen as soon as possible and not wait until just before the June 1, 2015, deadline.	This information has been added to the GHS subsection.

References

76 Fed. Reg. 44338–44339 [2011]. Notice of Public Meeting on the NIOSH Document Titled: Criteria for a Recommended Standard: Occupational Exposure to Diacetyl and 2,3-Pentanedione.

76 Fed. Reg. 64353 [2011]. Request for Public Comment on Draft Document: Criteria for a Recommended Standard: Occupational Exposure to Diacetyl and 2,3-Pentanedione; Extension of Comment Period.

78 Fed. Reg. 78363 [2013]. Notice of Request for Comments on Chapters 6 and 8 of the NIOSH document titled: Criteria for a Recommended Standard: Occupational Exposure to Diacetyl and 2,3-Pentanedione.

Akpinar-Elci M, Travis WD, Lynch DA, Kreiss K [2004]. Bronchiolitis obliterans syndrome in popcorn production plant workers. Eur Respir J *24*(2):298–302.

Bailer AJ, Stayner LT, Smith RJ, Kuempel ED, Prince MM [1997]. Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies. Risk Anal *17*(6):771–780.

Bellinger DC [2004]. What is an adverse effect? A possible resolution of clinical and epidemiological perspectives on neurobehavioral toxicity. Environ Res *95*(3):394–405.

Bock M [1975]. Minimax estimators of the mean of a multivariate normal distribution. Ann Stat *3*(1):209–218.

Budtz-Jørgensen E, Keiding N, Grandjean P [2001]. Benchmark dose calculation from epidemiological data. Biometrics *57*(3):698–706.

CDC (Centers for Disease Control and Prevention) [2002]. Fixed obstructive lung disease in workers at a microwave popcorn factory--Missouri, 2000–2002. MMWR *51*(16):345–347.

Chaisson NF, Kreiss K, Hnizdo E, Hakobyan A, Enright PL [2010]. Evaluation of methods to determine excessive decline of forced expiratory volume in one second in workers exposed to diacetyl-containing flavorings. J Occup Environ Med *52*(11):1119–1123.

Cox-Ganser J, Ganser G, Saito R, Hobbs G, Boylstein R, Hendricks W, Simmons M, Eide M, Kullman G, Piacitelli C [2011]. Correcting diacetyl concentrations from air samples collected with NIOSH method 2557. J Occup Environ Hyg *8*(2):59–70.

Crump KS [1995]. Calculation of benchmark doses from continuous data. Risk Anal *15*(1):79–89.

Ghanei M, Tazelaar HD, Chilosi M, Harandi AA, Peyman M, Akbari HM, Shamsaei H, Bahadori M, Aslani J, Mohammadi A [2008]. An international collaborative pathologic study of surgical lung biopsies from mustard gas-exposed patients. Respir Med *102*(6):825–830.

Gloede E, Cichocki JA, Baldino JB, Morris JB [2011]. A validated hybrid computational fluid dynamics-physiologically based pharmacokinetic model for respiratory tract vapor absorption in the human and rat and its application to inhalation dosimetry of diacetyl. Toxicol Sci *123*(1):231–246.

Hubbs AF, Cumpston AM, Goldsmith WT, Battelli LA, Kashon ML, Jackson MC, Frazer DG, Fedan JS, Goravanahally MP, Castranova V, Kreiss K, Willard PA, Friend S, Schwegler-Berry D, Fluharty KL, Sriram K

[2012]. Respiratory and olfactory cytotoxicity of inhaled 2,3-pentanedione in Sprague-Dawley rats. Am J Pathol *181*(3):829–844.

Israel L, Kim T, Prudhomme J, Bailey R, Harber P [2009]. Workplace spirometry: early detection benefits individuals, worker groups and employers. [https://www.thoracic.org/professionals/clinical-resources/clinical-cases/workplace-spirometry-early-detection-benefits-individuals,-worker-groups-and-employers.php]. Date accessed: October 2016.

Kanwal R, Kullman G, Fedan KB, Kreiss K [2011]. Occupational lung disease risk and exposure to butter-flavoring chemicals after implementation of controls at a microwave popcorn plant. Public Health Rep *126*(4):480–494.

Kanwal R, Kullman G, Piacitelli C, Boylstein R, Sahakian N, Martin S, Fedan K, Kreiss K [2006]. Evaluation of flavorings-related lung disease risk at six microwave popcorn plants. J Occup Environ Med *48*(2):149–157.

Kim T, Materna B, Prudhomme J, Fedan K, Enright PL, Sahakian N, Windham G, Kreiss K [2010]. Industry-wide medical surveillance of California flavoring manufacturing workers: cross-sectional results. Am J Ind Med *53*(9):857–865.

King MS, Eisenberg R, Newman JH, Tolle JJ, Harrell FE Jr., Nian H, Ninan M, Lambright ES, Sheller JR, Johnson JE, Miller RF [2011]. Constrictive bronchiolitis in soldiers returning from Iraq and Afghanistan. N Engl J Med *365*(3):222–230.

Kreiss K [2007]. Flavoring-related bronchiolitis obliterans. Curr Opin Allergy Clin Immunol *7*(2):162–167.

Kreiss K, Gomaa A, Kullman G, Fedan K, Simoes EJ, Enright PL [2002]. Clinical bronchiolitis obliterans in workers at a microwave-popcorn plant. N Engl J Med *347*(5):330–338.

Kullman G, Boylstein R, Jones W, Piacitelli C, Pendergrass S, Kreiss K [2005]. Characterization of respiratory exposures at a microwave popcorn plant with cases of bronchiolitis obliterans. J Occup Environ Hyg *2*(3):169–178.

Lockey JE, Hilbert TJ, Levin LP, Ryan PH, White KL, Borton EK, Rice CH, McKay RT, LeMasters GK [2009]. Airway obstruction related to diacetyl exposure at microwave popcorn production facilities. Eur Respir J *34*(1):63–71.

Maier A, Kohrman-Vincent M, Parker A, Haber LT [2010]. Evaluation of concentration–response options for diacetyl in support of occupational risk assessment. Regul Toxicol Pharmacol *58*(2):285–296.

Martyny JW, Van Dyke MV, Arbuckle S, Towle M, Rose CS [2008]. Diacetyl exposures in the flavor manufacturing industry. J Occup Environ Hyg *5*(11):679–688.

Morgan DL, Flake GP, Kirby PJ, Palmer SM [2008]. Respiratory toxicity of diacetyl in C57BL/6 mice. Toxicol Sci *103*(1):169–180.

Morgan DL, Jokinen MP, Price HC, Gwinn WM, Palmer SM, Flake GP [2012]. Bronchial and bronchiolar fibrosis in rats exposed to 2,3-pentanedione vapors: implications for bronchiolitis obliterans in humans. Toxicol Pathol *40*(3):448–465.

National Toxicology Program [2013a]. Testing status of 2,3-butanedione (diacetyl). [https://ntp.niehs.nih.gov/testing/status/agents/ts-m940009.html]. Date accessed: August 2013.

National Toxicology Program [2013b]. Testing status of acetoin. [http://ntp.niehs.nih.gov/?objectid=BDC958F4-123F-7908-7B566135E40DAE85]. Date accessed: August 2013.

NIOSH [1995]. NIOSH recommended exposure limit policy. In: NIOSH policy statements. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health.

NIOSH [2003a]. Hazard evaluation and technical assistance report: Agrilink Foods Popcorn Plant, Ridgway, IL. By Sahakian N, Choe K, Boylstein R, Schleiff P. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2002-0408-2915.

NIOSH [2003b]. Hazard evaluation and technical assistance report: Nebraska Popcorn, Clearwater, NE. By Kanwal R. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2002-0089.

NIOSH [2004a]. Hazard evaluation and technical assistance report: American Pop Corn Company, Sioux City, IA. By Kanwal R, Boylstein R, Piacitelli C. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2001-0474-2943.

NIOSH [2004b]. Hazard evaluation and technical assistance report: ConAgra Snack Foods, Marion, OH. By Kanwal R, Kullman G. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2003-0112-2949.

NIOSH [2004c]. NIOSH respirator selection logic. By Bollinger N. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2005-100.

NIOSH [2006]. Hazard evaluation and technical assistance report: Gilster-Mary Lee Corporation, Jasper, MO. By Kanwal R, Kullman G, Fedan K, Kreiss K. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2000-0401-2991.

NIOSH [2007]. Hazard evaluation and technical assistance report: Carmi Flavor and Fragrance Company, Inc., Commerce, CA. By Kanwal R, Kullman G. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2006-0303-3043.

NIOSH [2008a]. Evaluating occupational exposures and work practices at Agilex Flavors, Inc. formerly Key Essentials, Inc., Rancho Santa Margarita, CA. A technical assistance report to the California/Occupational Safety and Health Administration. By McKernan L, Dunn K. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2006-0361-2.

NIOSH [2008b]. Hazard evaluation and technical assistance report: Gold Coast Ingredients, Inc., Commerce, CA. By Bailey R, McKernan L, Dunn K, Sahakian N, Kreiss K. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2007-0033-3074.

NIOSH [2009]. Hazard evaluation and technical assistance report: Chr. Hansen, Inc., New Berlin, WI. By Sahakian N, Kullman G, Dunn K, Kanwal R. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2007-0327-3083.

NIOSH [2011a]. External review draft criteria for a recommended standard: occupational exposure to diacetyl and 2,3-pentanedione. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health.

NIOSH [2011b]. Hazard evaluation and technical assistance report: lung function (spirometry) testing in employees at a flavorings manufacturing plant -- Indiana. By Kreiss K, Piacitelli C, Cox-Ganser J. Morgantown, WV: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, NIOSH HETA Report No. 2008-0155-3131.

Park RM, Bowler RM, Roels HA [2009]. Exposure-response relationship and risk assessment for cognitive deficits in early welding-induced manganism. J Occup Environ Med *51*(10):1125–1136.

Rincon-Delgadillo MI, Lopez-Hernandez A, Wijaya I, Rankin SA [2012]. Diacetyl levels and volatile profiles of commercial starter distillates and selected dairy foods. J Dairy Sci *95*(3):1128–1139.

Sand S, von Rosen D, Eriksson P, Fredriksson A, Viberg H, Victorin K, Filipsson AF [2004]. Dose-response modeling and benchmark calculations from spontaneous behavior data on mice neonatally exposed to 2, 2', 4, 4', 5-pentabromodiphenyl ether. Toxicol Sci *81*(2):491–501.

van Rooy FG, Rooyackers JM, Prokop M, Houba R, Smit LA, Heederik DJ [2007]. Bronchiolitis obliterans syndrome in chemical workers producing diacetyl for food flavorings. Am J Respir Crit Care Med *176*(5):498–504.

van Rooy FG, Smit LA, Houba R, Zaat VA, Rooyackers JM, Heederik DJ [2009]. A cross-sectional study of lung function and respiratory symptoms among chemical workers producing diacetyl for food flavourings. Occup Environ Med *66*(2):105–110.

Wang ML, Petsonk EL [2004]. Repeated measures of FEV_1 over six to twelve months: what change is abnormal? J Occup Environ Med *46*(6):591–595.