Challenges in the study of causation of obesity

Thorkild I. A. Sørensen
Institute of Preventive Medicine, Copenhagen University Hospitals, Centre for Health and Society, Copenhagen DK1357K, Denmark

Use of the energy balance equation for understanding the causation of obesity is discussed. Its basis on the thermodynamic laws is expressed in mathematical models for body-weight changes. Only a very small net energy surplus per time unit constitutes the energy deposition during weight gain, making measurements of its components difficult. The physical laws provide exact quantitative relationships between energy intake, energy expenditure and deposition of energy, but cannot disentangle the initiating and driving forces of the energy imbalance, which may also be an active storage of fat in adipose tissue. These and various other limitations of the energy balance model warrant cautiousness in using the model in studies of obesity causation. Weight gain may be self-promoting and mathematical feedback models allowing estimation of such effects show that they are realistic. Predisposition and susceptibility should be distinguished, and susceptibility as a modifiable predisposition, the genetic and environmental contribution to predisposition and its usefulness as targets for prevention and treatment are discussed. Current progress in unravelling genetic predisposition, the complex genetically-determined mechanisms, the slower progress in unravelling the environmental influences, the different nature of genetic and environmental influences, the possible pathways of environmental influences and the environmental influences as mediators of genetic effects are addressed. The evidence behind the prevailing concept of the ‘obesogenic’ environment is critically analysed. Finally, particular opportunities for the identification of the causes of the obesity epidemic by detailed analysis of an observed irregular development of the epidemic over long time periods are presented, and evidence for predisposition as a result of postnatal environmental influences is inferred from these studies.

Causation of obesity: Energy balance: Obesogenic environment: Obesity epidemic

The study of the causation of obesity presents numerous challenges, several of which do not have clear solutions within the framework of currently-existing knowledge and available materials and methods. However, before considering the scientific challenges, there is a more cultural one. Generally, the prevailing opinion seems to be that the causes are well known and that the solution is simple; obesity is a problem emerging from gluttony and sloth, and if these factors are avoided, obesity will disappear. Discussion and debate from time to time raise some doubt and interest, but soon the prevailing opinion establishes itself again. It seems likely that this situation will not change until there is a much better understanding of the causation of obesity and the value of such knowledge in both the prevention and treatment of obesity has been demonstrated.

In the present paper some of the more prominent principal challenges, in the author’s view, will be presented and discussed. The discussion will be based primarily on theoretical considerations about the causation of obesity rather than inferences from the currently-available empirical evidence, which is rather fragmentary in this field and far from providing clear and useful knowledge of the causation of obesity. Concepts and components of the energy balance equation, the possibility of self-promoting weight...
gain, predisposition and susceptibility and the contention about the so-called obesogenic environment will be addressed, together with a suggestion of what may be a major challenge in the understanding of the causes of the obesity that constitute the obesity epidemic.

Energy balance equation

In steady-state the energy balance equation indicates that energy input (EI) equals energy output (EO), so that EI - EO = 0. A positive imbalance, EI > EO, implies that some energy is stored (ES), which is achieved by using energy to convert the surplus to tissue mass (termed Ec), i.e. EO = PAF × BMR. The steady-state balance can therefore be described as EI = PAF × BMR = 0 or EI = PAF × BMR, and the positive imbalance equation becomes EI - (PAF × BMR + Ec) = dES/dt. While taking into account that BMR is closely correlated with body weight for given body composition in terms of lean and fat tissue, this equation can be developed to a mathematical model for the relationship between changes in EI and in PAF and the changes in body weight, and can be used to predict change in body weight over time under various conditions\(^{(2)}\). Using known physiological data in the model, it is possible to delineate the permissible relationships. An example of the application of the model is that a man of 80 kg with a BMR of 7.8 MJ/d and a fixed PAF of 1.5 by adding 1 MJ/d to his food intake will gain about 10 kg in body weight during 2 years. The energetic value of the extra 10 kg is about 300 MJ if all fat, 40 MJ if all lean tissue and 235 MJ if 25% of the weight gain is lean and the rest is fat tissue. Thus, ES is much less than the cumulative increased EI, which is explained by the fact that the increase in body mass increases the energy requirements.

Very small energy imbalance during weight gain

Prediction of the greater energy expenditure due to the increase in energy needs of the added tissue depends on the proportionate increase in fat and lean body mass, although of course with a considerable individual variation\(^{(3)}\). It remains a challenge to measure the additional food intake (or reduced physical activity level) that contributes to the energy deposited as additional fat tissue. So far, the only effective way of estimating it is by measuring the amount of body mass added, converting it to energy and dividing this amount by the time needed to accumulate it, which will usually provide a value of \(<1\%\) EI (or EO) per unit time. The question may arise as to whether it makes sense to aim at measuring this very small difference between EI and EO. It follows from the physical laws of thermodynamics that there will be a difference and its magnitude can be estimated from the accumulated energy reflected in the weight gain. For the increase in ES to correspond to an increase in fat and lean mass in the usual relatively-fixed proportion there is of course a requirement that there is no internal net replacement of energy stores.

Initiating and driving forces of the energy imbalance

On the other hand, it is of considerable interest to identify the biological initiating and driving forces of the imbalance. It is worth noting that the energy balance equation as such is purely descriptive, defining the quantitative relationship between the physiological energy variables involved in weight gain. The mere fact that there is a difference between EI and EO exactly corresponding to deposited energy is not at all informative in relation to the causation of the process. The equation as such gives no hints about what could be the initiating and driving forces, and the equation does not put any restrictions on the options of which of the components could be involved. Thus, the process may be initiated and continuously driven by increased food intake, decreased energy expenditure as a result of reduced BMR or PAF (even though the weight gain by itself contributes to increase energy expenditure by the increase in BMR), as well as primarily an increased tendency to accumulate energy as fat in the adipose tissue.

Active storage of fat in adipose tissue

The latter possibility is particularly challenging in that it opposes the long-standing, but still prevailing, concept of the adipose tissue as a passive storage organ for surplus energy. It is now recognised that the adipose tissue is a biologically very active organ that is in intensive signal communication with the rest of the body. However, the notion that the fat storage function is just a reflection of the positive energy balance created by a primary difference between EI and EO remains an intuitively accepted idea, prevailing even among the majority of researchers of obesity who focus on adipose tissue functions\(^{(4)}\). Since obesity by definition is a condition with excess adipose tissue in which excess fat is stored as TAG, it seems paradoxical to overlook the fact that the problem of obesity primarily lies in the organ affected, the adipose tissue, which naturally would be considered first for any other organ-specific disease. It is easy to imagine that the development of obesity begins with a regulatory-driven tendency in the adipose tissue to store more fat by an increased formation of mature adipocytes, an increased uptake or formation (lipogenesis) of fatty acids or reduced release of fatty acids (lipolysis) or a combination of these processes. The energy needed for the processes as well as the energetic value of the stored fat would be derived from the overall energy source of the body; unless this process draws on energy stored elsewhere, it would of course instantaneously be reflected in a positive energy balance by correspondingly increased EI, reduced EO or both.
Monitoring of the initiation of energy imbalance

If it were possible to monitor the initiating process, it could be imagined that a disentangling of the primary changes in EI, EO or dES/dt might be possible. Assuming a starting point in steady-state, i.e. EI = EO, an exact monitoring of their level over time would in theory reveal whether EI begins to increase before EO increases in response to the weight gain (depending on course on any concomitant changes in PAF). If the process starts with a declining PAF, this step in theory should be observable as taking place before EI and later possibly EO starts to rise as a consequence of the weight gain. If the process starts with a simultaneous increase in EI and decline in PAF, a scenario that seems equally likely, this step might in theory also be observable in the initial phase. If the process starts with an accumulation of energy as fat by a replacement of energy from the non-fat compartment to fat, it might be possible to detect this step, but if it starts with a net accumulation of energy, i.e. a positive dES/dt, which instantly will be reflected in a corresponding difference between EI and EO, it may not be possible to differentiate this step from a primary increase in EI or a primary decrease in EO. In the real world such exact continuous monitoring covering the time period around the initiation phase appears impossible. When the process is ongoing following either of these initiation steps, there seems to be no way to identify which way it started, even when assuming sufficiently-precise measurement of all the components, and such measurements will just confirm what is already known, i.e. that the physical laws of nature still apply.

Complexity in causal model requirements

There may be a relevant asymmetry in complexity between the possibility that fat accumulation is initiated by disturbed processes in adipose tissue v. the possibility that initiation is by increased food intake and/or reduced physical activity. As follows from the energy balance equation, any explanation of obesity development based on a primary increase in food intake or a primary decrease in EO by reduced BMR or physical activity would need an assumption that compensation by the other components of the energy balance is either absent or insufficient, i.e. that an increased food intake is not accompanied by a corresponding increase in EO and that a decrease in EO is not accompanied by a corresponding decrease in food intake. This very important assumption must be fully integrated in any causal model of obesity development based on EI and EO. If this assumption is not made, then a causation model is postulated that is contradicted by the observations of body-weight history, both among non-obese and obese subjects. Thus, despite considerable fluctuations in EI and EO over days, weeks or months, most individuals keep their body weight stable within fairly narrow limits for long periods of time, often years, or with only a slight increase with aging. Such complex assumptions are not needed for the straightforward model of initiation by fat accumulation in adipose tissue. The consistent findings that diet-induced weight loss in obese subjects is almost always followed by regain in body weight during the next 5 years, is much easier to explain by the drive to accumulate fat in the adipose tissue than by a primary mismatch between combined regulation of EI and EO. Similarly, the constant finding that weight gain induced by overfeeding, contrary to what is seen in the weight gain that has produced obesity, is followed by weight loss down to the pre-experimental level, which is simpler to explain by the lack of an altered adipose tissue function than by combined re-adjustment of EI and EO.

Counter-regulation of energy balance like a buffer system

In view of the ability of body-weight regulation to cope with quite extensive fluctuations, it is a great challenge to find out why in some individuals the very small excess EI or reduced EO that constitutes the initial imbalance takes place and continues without activation of the otherwise very effective counter-regulation. Contrary to the experimental manipulations of EI and EO, there seems to be no external hindrances to the counter-regulation (see later discussion on the obesogenic environment). One possible mechanism that may be implicit in the arguments for a causal model based on a positive energy imbalance induced by a primary change in EI or EO is that the counter-regulatory system is working like a buffer system with boundaries of function that when broken lead to the excessive accumulation of body fat. The evidence for such a mechanism would require the identification of the boundaries. Such a model would require two stages of boundaries; one in which the body weight begins to change and one in which this change leads to a re-setting of the level of range within which body-weight regulation operates. As is clear from the overfeeding studies of lean individuals and the weight-loss studies in obese patients, a relatively short-term (weeks or months) induced great deviation in EI is in some subjects enough to elicit body-weight change, but re-setting of the range of body mass regulation rarely occurs, if ever, and it is the latter type of response that is needed to explain the development of obesity.

Evidence requirement for the initiating and driving forces

The role of the previously discussed arguments is only to challenge the conventional ideas that obesity development is caused by increased intake and/or decreased physical activity, not to disprove that it may be so. The arguments demonstrate the stringency of the evidence needed to disentangle the initiating and driving forces behind obesity development. In view of the physical links between the three components of the energy balance equation (EI, EO and dES/dt) it may even be argued that in most cases of common obesity it is very difficult, if at all possible, to identify which of the three components is the initiating and driving force. If the primary process is the fat accumulation in the adipose tissue, then the subsequent greater need for, and hence intake of, energy is supposedly mediated by an accompanying increase in appetite and/or reduction in physical activity. Observing an increased appetite or reduced physical activity during obesity development does not therefore allow unrestricted assignment of these disturbances as initiating or driving forces. Other types of
evidence may suggest such processes as primary, e.g. specific localised diseases, traumatic lesions or physiologically-demonstrated primary dysfunctions of the hypothalamic regulation of appetite by hunger and/or satiety. Enforced reduction in physical activity may not lead to increase in body weight (although possibly to an increase in intra-abdominal fat mass)\(^8\).

**Broadening the concept of causation of obesity**

The tight physical relationship between the components of the energy balance equation leads to the question: what is meant by causation of obesity. Thus, any conditions that need to be fulfilled for obesity to develop may be considered a cause of obesity. If a theoretical alteration in the processes regulating fat accumulation in adipose tissue cannot be met because of a blocking of the immediately associated demand for energy, either because of, for example, famine limiting EI or because of competing energy demands, e.g. increased BMR or increased physical activity, then this scenario illustrates the possibility that increasing EI or decreasing EO can be considered as causes of obesity development. Similarly, if the increasing need for energy because of the increasing BMR as weight is gained cannot be met because of restrictions in the availability of food, then this condition blocks further weight gain, and food availability may hence also be considered a cause. Rather than opposing the implicit consequences of the weight gain on energy needs by restricting food availability or enforcing an increase in EO, it may be a better solution to interfere with the initiation processes. However, the assumption that the initiation process can be blocked by restricting the availability of food for EI or increasing the EO by an enforced increase in PAF must take into consideration that the difference between the steady-state situation and the state after the initiation process is so small that it is not feasible to administer these restrictions without exaggeration. The consequence of exaggeration is the activation of counter-regulation, as seen in the experimental settings. To help in prevention and perhaps even treatment of obesity the great challenge must be the identification of the initiation processes regulating fat accumulation in adipose tissue.

**Interaction over time between causes and effects**

There have been a number of studies of determinants of weight gain, also investigated in large-scale long-term cohort studies, although few if any of the studies can draw conclusions about the full course of development from the non-obese state through to the obese state. With the measurement techniques available for both EI and energy expenditure (physical activity in particular) and for body-weight changes, it has so far not been possible to convincingly detect initiating differences in EI or energy expenditure and subsequent changes in body weight\(^9\text{–}16\). A major fundamental problem in such studies is that body weight tends to fluctuate; increases tend to be followed by decreases in body weight (as in the overfeeding studies) and decreases tend to be followed by increases in body weight (as in the weight-loss studies)\(^9\text{–}11\). To avoid bias for this reason, studies starting at any specific point in time will therefore have to take into account the possible effects of the preceding changes in weight on the determinants of subsequent changes in body weight. The risk of spurious inferences may be illustrated by the well-known phenomenon that weight loss is accompanied by a decline in BMR; the tendency to regain weight after weight loss will therefore implicitly make a low BMR a predictor of weight gain, as observed in one study\(^17\), even though it may not be involved in the process of regain. Similar plausible arguments may be put forward for the behavioural determinant of energy balance; food intake and physical activity at any point in time may be influenced, more or less deliberately, by the preceding trends in body weight and hence become predictors of the opposite trends in subsequent weight changes.

**Extreme distributions of energy balance components**

Many studies (including the author’s studies) have sought evidence for the role of food intake and physical activity in weight gain and obesity development by investigating the correlations between these exposures taken separately through a broad range of levels at one point in time and the subsequent change in body weight\(^9\text{–}16\). For such studies to be valid as tests of the role of these exposures in weight gain, a number of assumptions must be made that are not likely to be fulfilled. The first assumption is the straightforward one that measurements of food intake and physical activity have the same validity throughout the ranges, which is questionable because of possible level-dependent and correlated reporting bias. The second assumption, assuming the first one however is fulfilled, is that the more extreme the levels are the more likely is it that there is a concurrent energy imbalance, and the more extreme the levels are the greater is this imbalance. The third one is that the greater the imbalance, the less likely is it that the counter-regulatory mechanism can re-establish energy balance, possibly because an assumed buffering capacity in the counter-regulatory system is exceeded, as discussed earlier. The fourth one is that the more extreme the levels of food intake and physical activity, the less likely it is that what is observed for levels at the one point in time is not just transient fluctuations. Obviously, it is a series of assumptions that are of doubtful plausibility and very difficult or impossible to test. Even analysing the joint distributions of individual differences in levels or changes over time in food intake and physical activity and then looking at the extremes in the exposure matrix will still be subject to these assumptions. Adjusting one of the exposures for the other and thereby aiming at deriving a residual distribution supposedly quantitatively indicating the energy imbalance is a theoretically possible approach, but rests on several of the same assumptions, and it is particularly sensitive to level-dependent and correlated measurement errors. Most studies have failed to show any association between the behavioural exposures and subsequent changes in body weight. One study deserves particular attention because of the objective measurement of BMR (in Pima Indians), which was shown to be inversely associated with subsequent weight changes\(^17\). However,
for this association to be a valid demonstration of the importance of the position in the range of energy balance components, those with low BMR must not have lost weight before the BMR measurements, as discussed earlier. Furthermore, it is not clear why the energy balance regulation should work less efficiently the lower the BMR.

Limitations in lessons from manipulation of the energy imbalance

It has been frequently argued that manipulation of the EI and EO components of the energy balance equation, either experimentally or clinically, creating a positive or negative energy balance and subsequent changes in body weight is proof of the importance of altered EI or EO in the development of obesity. This argument is spurious for several reasons. The first reason is that such manipulation will have to comply with the physical law described by the energy balance equation, and hence the observations, showing that the law of nature is still valid, are completely uninformative in relation to the causes of obesity. The second reason is that the manipulations carried out by far exceed what can be estimated to be the magnitude of the energy balance disturbance reflecting the surplus storage of energy while obesity development is initiated. The third reason is that the manipulations have only been feasible for much shorter time periods than is required for obesity development, and hence will not be able to demonstrate that they will induce obesity in the long term. The fourth reason is, as mentioned earlier, that weight gain associated with overfeeding is subsequently associated with weight loss to the original level, and hence does not produce a condition that is similar to the obese state, in which weight loss is almost always followed by a regain in weight to the original level. These arguments obviously do not contradict the theoretical possibility that deliberate continued overfeeding over a long time period will eventually induce a condition similar to the obese state. The point to make is that the available evidence is not informative in relation to how common obesity develops.

Challenges in studying energy imbalance in obesity development

The straightforward implication of these considerations is that elucidation of the role of the primarily altered components in the energy balance equation in the development of obesity requires investigations of the development of obesity as it is initiated. For human subjects this type of investigation can only be carried out by studying populations of sufficient size and for a sufficient period of time to allow the observation that obesity development from the non-obese state has in fact taken place in an adequate number of subjects and that it does not take place in a sufficient number to serve as a comparison. Moreover, in these individuals sufficiently-accurate measurements of the components of the energy balance equation must be carried out before and during the initiation of the development of obesity to find out how the process starts. Until such studies become available claims about the role of changes in the energy balance components in the causation of obesity remain equivocal and can at best be extrapolations far beyond the currently-available evidence.

Self-promoting weight gain

The previously developed model for body-weight gain based on the energy balance equation, implies that an increase in EI or a reduction in EO by reducing physical activity by a fixed amount results in a weight gain that asymptotically reaches a new level, mainly because the increasing body mass induces an increase in BMR that dampens the tendency to weight gain, and eventually a new energy balance will be reached. However, a number of studies have shown that weight gain and obesity are associated with a subsequent decline in physical activity together with an increase in food intake to satisfy the increases in energy expenditure that results from increases in BMR, to the extent that the energy expended on physical activity is not reduced correspondingly. In this context it should be noted that the energy requirements for a given level of physical activity that moves the heavier parts of the body increase with increasing body weight, implying that less-physically-active obese subjects may still expend the same amount of energy on physical activity as more-active lean subjects. The body reactions to weight gain create a possible theoretical scenario in which the weight gain as such establishes conditions for continued weight gain, so that the weight gain may be considered to be self-promoting.

Mathematical feedback model

The process may be mathematically modelled as a feedback model that delineates the conditions for stable or unstable body weight. Using the previous model and plausible physiological values, the conditions for instability may be estimated; the results are quantitatively fully compatible with what may be considered realistic effects of weight gain. The instability may be considered a slight imprecision in the body reactions to weight gain; a little over-shooting in the increase in EI and/or in the reduction in physical activity as consequences of weight gain make the condition unstable, leading to further weight gain.

Quantitative effects of allowing feedback

The effects of the instability may be illustrated by the example presented earlier. In order to gain a certain amount of body weight with a fixed PAF of 1.5, e.g. approximately 10 kg in 2 years for a man with a starting body weight of 80 kg and a BMR of 7.8 MJ/d at the starting point, the requirement under the previous model is the addition of 1 MJ/d to the EI. Under the proposed feedback model, with a decline in PAF of 0.005/kg weight gain and an increase in EI of 0.15 MJ/d as a consequence of the weight gain, the same weight gain of about 10 kg will be achieved over 2 years by an initial 1% (approximately 0.1 MJ/d) increase in EI, i.e. one-tenth of the excess that results in the gain under the previous model. The course of weight gain will, however, be different. Where the previous model would result in an asymptotic levelling off at
the new level of body weight, with about half the gain achieved during the early part of the first two years, the feedback model will show an exponential rise with the major part of the weight gain during the second year.

Interpretation and value of the feedback model
Clearly, the two models do not preclude each other, but as mentioned by Christiansen et al.\(^{(19)}\), ‘allowing for the feedback may considerably reinforce the effects of even rather small primary excess intakes’. It is worth emphasising that the proposed feedback model, like the general energy balance equation, is only a mathematical description of the quantitative conditions under which weight gain may be self-promoting; it is not a proposition of a particular physiological feedback mechanism characterising those individuals who develop obesity \(v\), those who do not. As with the previous model, the feedback model will be extremely difficult to test empirically, if at all possible, because the predicted aberrations are far below the sensitivity of any available measurement tools applicable in the setting of the large-scale long-term population studies. The value of the model derives from the possibilities of showing that it is realistic, given observable physiological parameters, that weight gain could be self-promoting as a result of the slight instability in the body reaction to ongoing weight gain. Where the observational studies have so far failed to demonstrate that weight gain and eventual development of obesity are preceded by measurable excess EI or reduced physical activity, the feedback model shows that both increased EI and reduced physical activity may play a crucial role even within ranges that are even further away from what can be measured with any relevant accuracy. While in no way challenging the validity of the energy balance model, these options reinforce the challenges in using the energy balance model as a basis for the search for the causes of obesity.

Predisposition and susceptibility
It follows from these arguments that it may be difficult or even impossible to decide whether obesity development when initiated and ongoing is primarily driven by a slight excess EI, a slight reduction in energy expenditure or a slight tendency to increased accumulation of energy as fat in the adipose tissue, all three facilitating a positive energy balance. However, there are considerable individual differences in actual obesity development that most likely is attributable to pre-existing differences in predisposition to obesity development. It may be important to clarify what is meant by predisposition in this context. It seems operationally feasible to define it as variable characteristics, traits or features of the individuals in a particular population who under the same current environmental conditions are associated with an increased likelihood of continuous weight gain from the non-obese state to the obese state.

Susceptibility as modifiable effects of predisposition
If this predisposition is dependent on the environmental condition to which the subjects are exposed, the predisposition may be defined as an individual susceptibility to the particular environmental conditions that modify the differential individual likelihood of becoming obese. Since obesity may not develop under extreme environmental conditions in which there is no or very limited food available, predisposition to obesity may generally be defined as susceptibility. However, it may be useful to maintain the distinction between predisposition and susceptibility under non-famine conditions. This approach would allow maintenance of the distinction between predisposition to obesity that operates unmodified by the normal or commonly occurring variation in the non-famine environmental condition and predisposition that is modifiable and hence constitutes the susceptibility. This consideration is not just theoretical; the strategy for investigating the nature of the predisposition will probably benefit from applying a different study design and methods depending on the type of predisposition. Thus, identification and characterisation of a predisposition that is modifiable by the specific environmental conditions would require a study design and methods that allow the assessment of the effects of variation of these environmental conditions, either created by interventions or observed as they are occurring without interventions, on the relationship between the putative predisposition and the likelihood of obesity development. On the other hand, such a complicated study design will not be needed if the predisposition is not susceptibility, i.e. not modifiable by the environmental exposures within the commonly-occurring range.

Genetic- and environmentally-based predisposition
Predisposition may be established either because of differences in genetic constitution or because of previous environmental influences that have produced alterations in the individual that persist until it becomes manifest in obesity development. The currently-available evidence shows that both types of predisposition may operate for obesity\(^{(20)}\). The concept of genetic susceptibility then implies that carrying a particular gene set, compared with not carrying this particular gene set, leads to an increased likelihood of obesity development, the presence or magnitude of which depends on particular environmental conditions. Analogously, the concept of environmentally-based susceptibility implies that a particular environmental influence at one stage in life leads to an increased likelihood of obesity development, the presence or magnitude of which depends on later exposure to particular environmental conditions.

Targets for prevention and treatment
Obviously, all routes to the individual predisposition are of utmost importance and major challenges for future studies of the causation of obesity. A known genetic predisposition could lead to targets for prevention or treatment that operate by interfering with the gene function in the broadest sense, from gene expression through any regulatory, signalling or metabolic process that may determine the likelihood of later obesity development. The same principle applies to environmentally-based predisposition. For both
types the predisposition is also a susceptibility to particular environmental exposures later in life and the identification of these exposures is of similar importance as it creates an option to prevent obesity or prevent regain in weight after treatment of obesity by avoidance of these environmental exposures by those individuals who are susceptible to them. In the author’s view, comprehensive specific knowledge about predisposition to human obesity that has a proven value in translation to the preventive actions in the clinical and public health arena is currently quite far-off.

Unravelling the genetic predisposition

On the other hand, there are also grounds for optimism\(^{20,21}\). There exists a series of thoroughly-investigated rare syndromes with obesity as a distinct feature (the most well-known perhaps being the Prader-Willi syndrome) for which there is no doubt that they have a genetic variation as their basis. Several rare monogenic forms of obesity exist; the most clearly defined of these forms are those affecting the hypothalamic signalling pathways, with the defects in the leptin secretion from adipose tissue and in the leptin receptor in the hypothalamus as the first very inspiring and prominent examples, which were subsequently followed by unravelling of pathways involving the melanocortin-4 receptor (encoded by the \(MC4R\) gene) in the hypothalamus\(^{20,21}\). Multiple twin and several adoption studies have provided solid evidence that there is also a genetic predisposition to common obesity\(^{20}\). Application of modern explorative genomic technology has very recently revealed two genetic variations, one in the first intron of the \(FTO\) gene\(^{22}\) and one downstream of the \(MC4R\) gene\(^{23}\) that are closely associated with obesity and hence may contribute to the general predisposition to obesity. In other common chronic complex medical conditions, e.g. type 2 diabetes, very rapid progress has been made in recent years in the identification of multiple specific genetic variants that contribute to the predisposition\(^{24}\). A similar development is underway in relation to genetic predisposition to obesity, although there is a very long way to go since only a small percentage of the genetically determined variation in body weight is explained by the new genes.

Complex genetically-determined mechanisms

The mechanism by which the genetic variants found so far may lead to obesity is not clear\(^{25}\), but it is a major challenge and under intensive investigation. The assumption has been that the genetic variation constituting the predisposition is related in some way to subsequent variations in the structure or quantitative expression of the proteins encoded by the genes in the neighbourhood of the detected genetic variants. This approach means that the focus has been on the variation in the 2% of the entire genome, which hitherto has been considered ‘junk’ DNA, is indeed heavily involved in the cellular functions through many different forms of transcribed non-coding RNA, and the transcription may be in both directions of the genome (sense and anti-sense), increasing the potential information conveyed in the human genome from \(3 \times 10^9\) to \(6 \times 10^9\) base pairs\(^{27}\). These new even more challenging dimensions must be included in the search for the mechanisms or pathways by which the genetic variation results in variation in predisposition to obesity.

Less progress in unravelling the environmental influences

A corresponding development in acquiring new knowledge has not (yet) taken off at the same rate in the search for the environmental influences that may contribute to predisposition to obesity, or in the search for specific environmental factors that later in life may interact with the genetic- or environmentally-based predisposition that may make the individuals susceptible to such exposures later in life. There are several obvious reasons why progress has been slower than in genetics. The genome is assumed to be a persistent structure within each individual over time, in fact from conception to death, even though there are possibilities for changes in the genome during a life course (errors in cellular replications, mutations). However, such changes are likely to affect the genome only in single cells or subsets of cells, as happens in cancer development. Thus, with the advances in biotechnology, especially in the tools used to investigate the composition of the genome, the search for the genes responsible for the genetic predisposition to obesity has a massive advantage compared with the search for specific environmental factors.

Different nature of genetic and environmental influences

By definition environmental factors are effective at a particular time and place for each individual and hence are potentially subject to continuous changes over time. Moreover, the opportunities to observe and quantify environmental exposures, hitherto presumed to be of importance, are very limited compared with the opportunities to investigate the genome. The effects of the environment may exhibit continuous linear or non-linear types of dose–response effects, may have thresholds, may have a time-lag or cumulative effects and may interact and be correlated with other environmental exposures, changing over time. Whereas the geneticists can base their studies on snapshots of the genome at any time during the subject’s life until a selective morbidity and mortality related to the genomic profile is established, the ‘environmentalist’ will have to carry out the investigations of the exposure that may contribute to the predisposition at one defined stage in life and then wait until this influence may become evident in obesity development at a different rate from that observed in other subjects not exposed to the same environmental influences at that particular stage of life. Usually, this requirement will imply that the waiting time is many more years than are possible in any realistic duration of a research project. As a result researchers interested in these aspects will have to rely on previous data for a large
number of subjects, which would have been collected long before they themselves could be involved and before the appropriate current technology was available or even developed. However, despite these difficulties some progress has been made(28), although factors of sufficient specificity to be the basis for detailed investigation of the mechanisms or the basis for attempts at translation into preventive actions in public health settings are still lacking.

Pathways for environmental influences
Understanding the mechanisms by which early-life environmental influences may contribute to a predisposition to obesity that may not emerge until much later in life is another important challenge, because such knowledge may define an appropriate target and lead to more-precisely-administered preventive actions(29). To be a predisposition some persistent changes must occur in the subject such that their influence would assume a similar status in the mind or body of the subjects to that of a genetic variation. Persisting modification of structure–function relationships in the developing brain, especially in the hypothalamic centres and the connectivity with other parts of the brain are possible and interesting options. Another option under current scrutiny is epigenetic mechanisms, based on stable methylations of the genome in particular sites and histone formations that may influence the gene expression activity later in life. In view of the dynamics of these regulatory features of the genome and also the new previously mentioned complexity of the function of the genome (copy number variants, non-coding RNA, anti-sense transcription), it is a major challenge to identify the selective stable alterations of the genome that could be the basis for environmental influences with long-term effects(29).

Environmental influences as mediators of genetic effects
There is another and important asymmetry in the search for genetic v. environmental factors that may contribute to the difficulties in identifying proper environmental determinants, which applies to both environmental exposure influencing the predisposition and environmental exposures that may interact with the genetically-determined predisposition. Where the genome can be investigated as such, it can be shown that what are considered to be and investigated as environmental exposures may in fact reflect genetic influences. Thus, the microenvironment created for the foetus by the pregnant woman and by the parents of the newborn may very well be determined by their own genetic constitution, part of which is also directly transmitted to the child(30). Later in life the particular genetic constitution of the individual may lead them intentionally to express a particular behaviour and/or to be exposed to a particular environment, which then operates as a mediator of the genetic effects rather than an additional exposure alongside the genetic factors(31–33). There is also some evidence emerging from animal experiments that epigenetic alterations in the genome in one generation may also affect the germ cells and thereby be transmitted as if it were truly genetic variation in the subsequent generations(29). It is therefore frequently difficult, and often impossible, to disentangle accurately what are independent environmental factors in obesity development.

Obesogenic environment
The obesity epidemic has developed rapidly during the last decades primarily in the Western world, but more recently also as a pandemic in most other populations in the world except those with a continuous shortage of food and frequent famines. It has led to the widespread use of the term ‘the obesogenic environment’, which is intended to indicate that the epidemic is caused by particular changes in the environment(34–39). There is, of course, no doubt that some environmental changes must have induced the increases in the occurrence of obesity, since it is argued, based on current knowledge, that variation in the human genome at the population level cannot occur so rapidly. The crucial questions are of course which changes have taken place and how have they affected the development of obesity, in particular whether the changes affect the predisposition, whether they interact with the predisposition and whether they are directly involved in the process of creating the positive energy balance while obesity develops. This knowledge may not be as well established as is frequently assumed by lay individuals, by healthcare professionals who are involved with the problems of obesity and even among scientists working in the obesity field.

Obesogenic environment as a prevailing concept
The concept has prevailed until now in many settings including, for example, in WHO(38). In accordance with what might be considered a naive interpretation of the conditions for energy balance regulation, as discussed earlier, the concept appears both intuitively appealing and self-evident. Translation to the public health sector should therefore not pose any problems; the given tasks are to change the environment so that it leads to reduced EI and increased levels of activity. The proposed changes in the environment are of such obvious expected benefits to the population that it is very difficult to maintain any critical position. However, it has been difficult to obtain further scientifically-convincing evidence in support of the broad concept of the ‘obesogenic environment’ as the cause of the pandemic of obesity. A major additional problem is the complete lack of investigated explanations related to the energy balance component of the considerable heterogeneity in levels and rates of the changes in obesity prevalence both between and within populations. This problem in itself does not show that the concept is wrong, but rather leaves the option that there might still be other explanations of the pandemic that perhaps are even more convincing, more specific and translatable to intervention targets and therefore even more appealing from a public health point of view.

Where is the evidence for the obesogenic environment
In a recent systematic review of the literature addressing associations between environmental factors, energy and fat intakes among adults the question was raised: ‘is there
In response, the authors claim that 'It is too premature to conclude whether or not environmental factors play a role in obesogenic and unhealthy dietary intakes. More studies need to examine associations with those environmental factors thought to contribute to obesogenic environments. There needs to be more development in theories that conceptualise the relationship between environmental factors and dietary intakes' (40). Another review of the literature has commented that 'Reduced physical activity, particularly from reduced school-based physical education, and specific food manufacturing and marketing practices (e.g. vending machines in schools, increased portion size, increased availability of fast-food, use of high-fructose corn syrup (HFCS)) comprise the Big Two explanations proffered for the obesity epidemic and are frequently cited as targets of potential public health interventions. We do not intend to imply that the Big Two are not salient contributors to the epidemic. Rather, we offer that the evidence of their role as primary players in producing the epidemic (as well as the evidence supporting their potential ability to reverse the trend if manipulated) is both equivocal and largely circumstantial – that is, the hypothesised effects are underdetermined by the data. Data rarely, if ever, stem from randomised controlled trials of the effects in population settings and in many cases do not even include a consistently supportive body of individual-level epidemiologic studies. The arguments for the effects of each subcomponent tend to rely heavily (although not exclusively) on presumed mechanisms of action and ecological studies in which associations between the putative factor and obesity rates are shown at the aggregate population level across times or geographic locations (41). A recent study of secular trends and geographical differences in objectively-measured physical-activity energy expenditure has found no trends or geographical differences matching what otherwise would be expected from the occurrence of obesity in the populations assessed (42).

**Exploring alternatives to the obesogenic environment**

Subsequent to the search for the supporting evidence for the ‘Big Two’, the previously mentioned review has explored ‘putative contributors to the secular increase in obesity: exploring the roads less travelled’ (41). The authors state: ‘Subsequently, we delineate the evidence for 10 other putative factors for which the evidence is also circumstantial but in many cases, at least equally compelling.’ The following factors were identified and the concordance in time trends with the obesity epidemic among adults in the USA were analysed: mean age of mothers at first birth; antidepressant prescribing (UK data); prevalence among households of air-conditioning; average internal home temperature; concentrations of polybrominated diphenyl ethers in the breast milk (Swedish women); proportion of the US adult population that is Hispanic and/or between 35 and 55 years of age; time spent awake; non-smoker prevalence. It was concluded that ‘undue attention has been devoted to reduced physical activity and food marketing practices as postulated causes for the epidemic, yielding neglect of other plausible mechanisms’ (41). All these other possible causes of obesity were found to exhibit increases during the three decades following 1970, although for most of them there remained doubt as to whether the increase started earlier than the rise in prevalence of obesity as they should have done if they were true causes of the obesity epidemic.

**Developmental phases in the obesity epidemic**

Usually, time-trend studies of the occurrence of diseases are not particularly useful in the search for causes of the disease, mainly because of the risk of the so-called ecological fallacies; at group levels there appears to be a parallel development in the occurrence of the putative cause and of the disease, but when analysed at the individual level there is no, or even the opposite, association. However, when the time trends are non-linear there are opportunities to narrow down the evidence for the possible association between the putative cause and the disease. Thus, if the occurrence of the disease shows the same irregularity as the presence of the putative cause and does so with an appropriate time lag it is more likely that there may be a cause–effect relationship between the two. If there is no concordance in these time trends, it may be concluded that the putative cause cannot play the role as the main responsible contributor to the change in prevalence that has hypothetically been assigned to it. As an example, the time trends in the incidence of lung cancer closely follow the irregularity of the prevalence of heavy cigarette smokers in the population with about a 15-year delay, strongly supporting the role of cigarette smoking as a main cause of lung cancer in addition to the findings of individual associations in cohort studies.

**The obesity epidemic in children and young adults in Copenhagen**

The development of the obesity epidemic among children and young adult men in the Copenhagen area since the interwar period has been investigated in detail (43–55). It was observed that the epidemic among the young men has developed in phases (43,45,47,49,53), the first one starting abruptly after about two decades of low stable prevalence in the birth cohort born in early 1940s and continuing for about a decade up to an approximately 10-fold higher prevalence of obesity. It then levelled off and was stable at this higher level for about two decades. A second even steeper rise in the prevalence began in the birth cohorts from 1970 onwards, one generation after the first increase. These four phases were also already seen in schoolchildren when they started school at ages 6–7 years and later, and the changes from one phase to another occurred in the same birth cohorts as it did for the young men (51–53,55). Even within the rather narrow age range of 18–24 years, during which the young men were examined when attending draft boards, the first rise in prevalence was clearly linked to birth cohorts, beginning abruptly in 1942 (43). Although there is a distinct tracking relationship between BMI throughout its range and the likelihood of being obese later in life (45,46), this development is not reflected in a
corresponding upward drift of the entire BMI distribution\(^{(44,47,49)}\). The median is almost constant throughout the time period in both the schoolchildren and the young men\(^{(51,55)}\). Commingling analysis of the distributions for the young men suggests that the increase in prevalence of obesity was associated with an increase in subjects in an upper of two otherwise positional stable distributions\(^{(48)}\). Similarly, the birth weight, which also shows a clear monotonic relationship with later BMI and risk of obesity\(^{(50,52)}\), does not show any corresponding trends in the overall distribution\(^{(52)}\). The four phases of changes in obesity prevalence are independent of the development of the economic prosperity of the country\(^{(55)}\); thus, the first rise in prevalence occurred before the economic growth began, the prevalence was stable when the economic growth accelerated, and when the second rise in prevalence of obesity began, there were no preceding changes in the rate of economic development. It is intriguing that prevalence data from UK\(^{(37)}\), USA\(^{(41)}\) and Australia\(^{(56)}\) also show phases in the development that look similar to those observed in Denmark, although the data are not as detailed and not analysed in the same way.

**Evidence for predisposition due to postnatal environmental influences**

It is suggested that the most likely interpretation of these observations taken together is that there has been a change in the environmental conditions that contribute to a crucial predisposition to later development of obesity, established some time before the age of 6–7 years as the influence on the epidemic is already fully evident at that age. Since it takes time, probably several years, to develop the obesity present at age 6–7 years, it is assumed that the influence operates in the early postnatal period or before birth. The finding that there were no notable corresponding changes in the birth-weight distribution over time and no change in its relationship with risk of later obesity during the four distinct phases of the evolvement of the epidemic suggests that the critical period is in the early postnatal life. On the other hand, it cannot be excluded that prenatal influences that are not mediated through changes in birth weight may operate, which is illustrated by increased risk of obesity in low-birth-weight children if the mothers had been smoking during pregnancy and by the increased risk associated with excessive gestational weight gain that cannot be explained by the increased birth weight\(^{(28)}\). During the subsequent period there has been an increase in birth weight\(^{(54)}\), but its implications for later occurrence of obesity remains to be seen. So far, there have been no specific indications of what has changed in the environmental conditions in early life. The perspective is of course that if what has changed can be successfully identified, possibly with help from professional historians, it means that it should also be possible to reverse these conditions, and hence remove the environmental predisposition that appears crucial for the development of the epidemic. If such intervention requires particular actions of the young parents on behalf of their newborn (or yet unborn) child, it may be assumed that the basis for such preventive action is better in this situation than in any other period in life for both the parents and their child. For these reasons, the greatest of the challenges is in identifying what in the perinatal period has caused the changes in the rate of the development of the obesity epidemic.

**Conclusion**

While acknowledging the unchallenged validity of the thermodynamic laws of the energy balance model as a basis for the investigation and understanding of the causation of obesity, a critical analysis shows that it has little usefulness. The finding that body reactions to weight gain may create conditions for self-promoting weight gain make it even more difficult to base the search for the causes of obesity on these laws. Investigations into the predisposition to obesity as a result of genetic variation or exposures to environmental influences with long-standing effects may be the route to finding targets for prevention and treatment of obesity. In particular the predisposition that constitutes a susceptibility to identifiable and modifiable environmental exposures later in life will be of relevance, allowing counteractions of the predisposition later on. While the concept of the so-called ‘obesogenic’ environment has a clear generic truth, its contents need to be studied in much more depth in terms of its role in causing the obesity epidemic before it can be used in obesity prevention and treatment. Detailed analysis of the development of the obesity epidemic may direct the attention particularly to the early postnatal environment.

**Acknowledgements**

The author declares no conflict of interest. Current industrial collaborations include metabolomic studies with Nestlé Research Centre, Lausanne, Switzerland, and genomic studies together with DSM, Heerlen, The Netherlands. This work is based on research in the Danish Obesity Research Centre (DanORC), supported by The Danish Council for Strategic Research (grant no. 2101-06-0005), and two integrated projects of the EU Sixth Framework Programme for Research and Technological Development: Diet, Obesity and Genes – Targeting the Obesity Problem: Seeking New Insights and Routes to Prevention (DIJENES; contract no. FOOD-CT-2005-513946), and Hepatic and Adipose Tissue and Functions in the Metabolic Syndrome (HEPADIP; contract no. LSHM-CT-2005-018734).

I thank all my many junior and senior collaborators, with whom over the years I have discussed the issues of this paper. I am particularly grateful to the mathematician, Edmund Christiansen, who sadly drowned in October 2007 in an attempt to rescue his wife from drowning during swimming from a beach in Sicily.

**References**


