Diurnal variation of circulating interleukin-6 in humans: a meta-analysis

Gustav Nilsonne¹,²*, Mats Lekander¹,², Torbjörn Åkerstedt¹,², John Axelsson¹,², Michael Ingre¹

¹ Stockholm University, Stress Research Institute, Stockholm, Sweden
² Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden

* gustav.nilsonne@ki.se

Abstract

The pleiotropic cytokine interleukin-6 (IL-6) has been proposed to contribute to circadian regulation of sleepiness by increasing in the blood at night to signal for sleepiness. Earlier studies have reported diurnal variations of IL-6, but phase estimates are conflicting. We have therefore performed a meta-analysis on the diurnal variation of circulating IL-6. Studies were included if they reported circulating levels of IL-6 recorded at least twice within 24 hours in the same individual. A systematic search resulted in the inclusion of 43 studies with 56 datasets, for a total of 1100 participants. Individual participant data were available from 4 datasets with a total of 56 participants. Mixed-effects meta-regression modelling confirmed that IL-6 varied across the day, the most conspicuous effect being a trough in the morning. These results stand in contrast to earlier findings of a peak in the evening or night, and suggest that diurnal variation should be taken into account in order to avoid confounding in studies of IL-6 in plasma or serum.

Introduction

Sleepiness is regulated in humans by two main processes: the circadian process, which makes us sleepier in the night, and the homeostatic process, which causes sleepiness to increase with time awake [1]. It has been proposed that interleukin-6, a pleiotropic cytokine, participates in circadian sleepiness regulation by increasing at night in the blood and inducing sleepiness through signalling in the brain [2-6]. Early studies of diurnal variation of IL-6 in humans found a peak in the night-time [7,8], and it is this observational relationship that forms the main line of evidence for a regulatory effect of circulating IL-6 on sleepiness. However, further studies have since found peaks at different times of the day or have found no peaks at all. Fig 1 shows locations of peaks and troughs that have been estimated in the literature so far. Notably, estimates have ranged quite widely. Nonetheless, the general impression of these earlier claims is consistent with an increase of IL-6 levels in the night-time.

One previous meta-analysis of IL-6 and time of day has been reported [9] (and published again in [10] and [11]). This meta-analysis focused on the diurnal variation of interleukin-6 in rheumatoid arthritis, but also included an estimate for healthy control participants from 11 studies. Data inclusion procedures were informal; no systematic method for identifying and including data was reported. The main finding in healthy
participants was an increase if IL-6 from the evening, continuing during the night, followed by a drop in the morning. The pattern in patients with rheumatoid arthritis was similar, but with a more pronounced peak in the early morning before levels started to fall.

Thus, the observational relationship between IL-6 and time of day in healthy humans has important implications for the theoretical understanding of immune-brain interactions in sleepiness regulation, but there is no consensus on estimates of phase. Therefore, we have performed a meta-analysis, aiming to investigate the diurnal variation of IL-6 in the blood.

**Fig 1. Estimates of phase reported in earlier literature.** Every count represents one claim of having located a peak (box above time-line) or a trough (box below time-line) in a dataset. Blue: Studies included in quantitative review. Green: Studies not included in quantitative review. Orange: Meta-analysis. Review papers are not included.

**Materials and Methods**

**Literature search and data acquisition**

The PubMed database was searched using the terms "interleukin-6 AND (sleep OR diurnal OR circadian)" and the limit "human". The search was last updated on 2016-01-03. Records were reviewed by one investigator (GN). Studies were included if they reported IL-6 in plasma or serum from healthy participants with a time-course including two or more time-points within 24 hours. Fig 2 shows a flowchart of data inclusion. Table 1 shows characteristics of included studies. Table 2 lists studies that fulfilled inclusion criteria but which could nonetheless not be included. The most common reason was that data could not be estimated (k = 25). Of these 25 studies, 7 reported that data were largely or entirely under the assay detection limit. In the remaining cases, data could not be estimated because they were given as a difference score (k = 5), because they were not shown (k = 4), because time of day was not given (k = 3), or for other reasons, specified in table 2 (k = 6). Additionally, seven studies were excluded due to duplicate publication of data, and four studies were excluded because the reported levels of IL-6 were very high and therefore judged not to represent levels consistent with physiological regulation or variation in healthy humans. Of these studies, one reported one participant, whose IL-6 levels increased ten-fold after venous catheterization [72], and we judged that this change was not representative of diurnal variation. Another study reported ten participants with mean plasma IL-6 levels of about 10-30 pg/ml over the course of two days [73]. This is about ten times higher than expected for healthy participants, raising questions about the validity of the absolute values. We judged that these measures may not accurately reflect the target outcome, and that they would unduly influence the regression model on account of the very high values. They were therefore excluded. Finally, two studies [83,84] reported 40 and 60 participants, possibly with overlapping samples, both with mean values of about 35 and 55 pg/ml at 09:00 and 02:00, respectively. Because these levels were so much higher than expected for healthy participants, these studies were also excluded. Thirteen studies were included from sources other than the PubMed results. These were found mostly because they were cited by papers identified in the literature search, and in some cases because they were known to us on beforehand.

Data were estimated from published tables or from graphs using GetData Graph Digitizer, version 2.25.0.25 (getdata-graph-digitizer.com). Error bars were
Fig 2. Data inclusion. Some of the 43 included studies contained more than one dataset. The final number of datasets was 56 (see table 1).

assumed to represent standard errors unless otherwise indicated. Time of day was coded, as well as sleep or wake, time asleep, and time awake. For studies reporting participants kept in the lab overnight, data obtained at the time-point when the lights-out period began were coded as awake and data from the time-point when the lights-out period ended were coded as asleep. When times for falling asleep and waking up were not recorded or not reported, we assumed that they were 23:00 and 07:00. When applicable, time from catheter insertion was also coded. Unless otherwise specified, serial sampling with more than two samples within the same 24-hour period was assumed to have been performed using an indwelling catheter inserted at the first sampling time point. Data recorded during sleep deprivation were not included. IL-6 data were ln-transformed to better approximate a normal distribution. For datasets where individual participant data were not available, transformation was performed as described in [12].

Individual participant data were available from 4 datasets, which were coded separately (see table 1). In these datasets, data points below assay detection limits (meaning lowest known point of assay linear range) were conservatively re-coded to the value of the detection limit. In Sothern 1995 [13], 7 values out of 88 (8%), ranging from 0.5 to 0.96 pg/ml, were re-coded to 1 pg/ml. In Karshikoff 2015 [14], 23 values out of 83 (28%), ranging from 0.01 to 0.88 pg/ml, were re-coded to 0.9 pg/ml.

Ethical approval was not required. The study protocol was not registered. All data and the full analysis code are freely available at https://github.com/GNilsonne/IL6_diurnal.

Meta-analysis

To investigate the diurnal time course of IL-6 in plasma and possible moderator variables, we used hierarchical mixed-effects models. This approach allows for more complex model fitting and is expected to have higher statistical power compared to fitting models separately in each dataset and then analysing summary measures such as acrophase and amplitude. Diurnal variation was investigated by fitting cosinor functions with periods of 24, 12, and 6 hours. Time from catheterisation was included with a random slope for each data set in order to account for the proposed effect that catheterization induces higher values in blood drawn from the catheter [15,16]. Effects of sleep were investigated exploratively with a binary factor for sleep/wake, as well as time asleep and time awake. Datasets were weighted by the number of participants multiplied by the square root of the number of time-points in each study. Models were compared using likelihood ratio tests. Analyses were performed using R version 3.2.0 [17] with the nlme package [18].

Results

Diurnal variation of IL-6

First, we fitted a null model including only time from catheterization and a random intercept for each dataset. Fig 3 shows residuals after these effects have been accounted for, suggesting that there remains variation to explain. The distribution of these residuals suggests a morning trough in IL-6 levels (Fig 3). Next, we compared a model with a 24 h cosinor function to the null model. The 24 h cosinor model fit better (log likelihood -483.5 vs -528.8, p < 0.0001, Fig 4). We then added another cosinor function with 12 h period. We did this for two reasons. The first reason was that addition of
shorter periods allows a better estimation of non-sinusoidal effects, albeit at a cost of
higher risk of overfitting. The second reason was that 12 h periods have been proposed
by earlier investigators [19], and we considered that these claims should be tested. The
model with both 24 h and 12 h cosinor functions fit better than the model with only the
24 h period (log likelihood -460.1 vs -483.5, \( p < 0.0001 \), Fig 4). Next, we exploratively
investigated the addition of yet another cosinor with a 6 h period, but that did not
improve model fit (log likelihood -457.7 vs -460.1, \( p = 0.09 \), prediction not shown).

**Fig 3. Residuals from null model.** Data points sized by regression weight. In the
null model, a random intercept for each study and a linear effect of time from
catheterization have been included. Therefore, these residuals show the putative diurnal
variation to be modeled. To explore this variation, we fitted a weighted LOESS curve
(red line). This curve shows a trough in the morning. Note that the shape of the
LOESS curve depends on the smoothing parameter. It is therefore possible to generate
different LOESS curves from the same data, and not all of them show a peak in the
early afternoon. The LOESS curve was fitted on three repeated days of the same data,
and the curve for the second day shown, to ensure that the estimates would meet at
00:00 and 24:00. Time from 22:00 to 07:00 is shaded to indicate the night. Left: All
data points shown. Right: Y axis range restricted to increase resolution.

**Fig 4. Predicted diurnal time courses from meta-regression models.** Left:
Best-fitting model including cosinor functions with 24 and 12 h periods. Right: Model
including only 24 h period. Time from 22:00 to 07:00 is shaded to indicate the night.

Attempting to disentangle diurnal variation from effects of sleep, we investigated the
addition of model effects for sleep (asleep/awake), time asleep (hours since sleep onset),
and time awake (hours since wake onset). Starting with the best-fitting model including
24 and 12 h periods, we found that the addition of sleep, time asleep and time awake, or
all three variables, did not improve model fit (log likelihoods -459.9, -459.2, and -459.0,
respectively, vs -460.1, with \( p \) values 0.80, 0.36, and 0.36). Finally, we investigated the
addition of sleep, time asleep and time awake, and all three variables, to the model with
24 h period. When comparing these models to the best-fitting model with 24 and 12 h
periods, we saw worse fit (log likelihoods -475.5, -471.5, and -471.5, vs -460.1, with \( p \)
values \( \leq 0.0001 \)). When comparing to the model with 24 h period only, we found that
the addition of sleep, time asleep and time awake, or all three variables, yielded
better-fitting models (log likelihoods -475.5, -471.5, and -471.5, vs -483.5, \( p \) values
0.0003, < 0.0001, and < 0.0001).

The best-fitting model with 24 and 12 h periods had a conspicuous trough between
09:00 and 10:00 in the morning and a second less pronounced trough close to 22:00,
and two peaks located close to 17:00 and 02:00 (Fig 4). Since diurnal rhythms are commonly
investigated using cosinor functions with 24 h periods, we show predictions from this
simpler model too (Fig 4). This model estimated bathyphase (lowest point) at 08:05
and acrophase (highest point) at 20:05, with an amplitude of 0.166. All the included
datasets, with predictions from the best-fitting model, are shown in figures 5 and 6.
Individual participant data are shown in figure 7, for those four studies from which
individual participant data were available.

**Assessment of risk of bias**

Our literature review identified 36 eligible studies which could not be included. Not
counting participants with duplicate data, these studies reported 468 participants,
compared to the 1100 included in our meta-analysis. Of the ineligible studies, seven,
Fig 5. Data and fitted time courses, showing the first 30 out of 56 datasets. Data are shown as estimated from original publications, with error bars showing standard deviations. Y axes show ln IL-6 (pg/ml) throughout. Hours are in chronological time where 1 is 01:00 on the first day. Red lines show predictions from the best-fitting model.

Fig 6. Data and fitted time courses, showing the last 26 out of 56 datasets. Data are shown as estimated from original publications, with error bars showing standard deviations. Y axes show ln IL-6 (pg/ml) throughout. Hours are in chronological time where 1 is 01:00 on the first day. Red lines show predictions from the best-fitting model.

Fig 7. Individual participant data. Individual participant data were available from four datasets and are shown here mainly for the purpose of illustrating the high degree of variability within individuals. To illustrate summary effects within each dataset, thick lines show loess functions fitted to each dataset. Time from 22:00 to 07:00 is shaded to indicate the night. For Lekander 2013, two sets of measurements, made with a few days’ interval, have been plotted over the same time course, in red and blue respectively.

most of which were published in the 1990s, were unable to find reliable data because most values were close to or below assay detection limits (Table 2). From the point of view of bias, this is not a major problem since assay detection limits were unrelated to time of day for sampling. More concerning are data that were not reported because no effect was found. Only one paper explicitly stated that data were not shown for this reason [66]. Additionally, an unknown number of studies with eligible data have never been published. It is likely that unreported studies were more likely not to have found significant effects. However, with regard to estimating diurnal phase, we suspect that earlier investigators have been happy to report effects regardless of the location of peaks and troughs, reflected in the wide variety of published estimates (Fig 1). Therefore, even though the first reports of diurnal effects reported peaks in the night, we suspect that the studies included here were not strongly biased towards reporting effects at any particular time of day. Furthermore, to the extent that earlier reports may have been influenced by a prevailing theory, the effect most often referred to is a night-time peak. Since our meta-analysis did not find a night-time peak, we think it is unlikely that a bias in favor of this particular effect will have had a major influence on our model estimates.

Included datasets comprise both data from studies that were designed to measure diurnal or circadian variation, and data from studies that incidentally happened to fulfill our inclusion criteria. The latter group generally had fewer time points for measurements. Since this meta-analysis uses mixed-effects meta-regression by time of day, and is hence not based on single summary measures of included studies, it is not possible to investigate heterogeneity and bias by usual means such as a funnel plot. The results did not strongly depend on any single study. The most influential study was Mehra et al. [44], with a total regression weight of 11.19% over five different datasets.

Since this analysis concerns an observational relationship, confounding from a variety of sources cannot be ruled out. Light exposure and photoperiod were not controlled nor recorded except in laboratory studies, and season or time of the year were not reported frequently enough to justify inclusion in coding. Similarly, physical activity can affect levels of IL-6, but any instructions to participants about physical activity, and measures of their behavior, were for the most part not reported in included studies. Our assumption that participants in datasets not specifying sleep times on average slept 23:00-07:00 would limit the ability to estimate effects of sleep, but that is beyond the
scope of this paper, and as the sleep variable was not included in the best-fitting model, the assumption only indirectly affects the risk of bias in diurnal variation. Since the sleep variable made very little difference when added to the best-fitting model, it is unlikely that a different assumption would lead to a different result.

Based on the above considerations, we judge that the risk of bias due to selective publishing and data inclusion is probably moderate to low.

Discussion

Our meta-analysis confirmed that circulating IL-6 shows diurnal variation. The most marked effect was a morning trough. The best-fitting model included a 24 h and a 12 h cosinor component. The LOESS curve shown in figure 3 suggests a relatively flat curve from the afternoon to the late night, and the better fit obtained by adding the 12 h component may reflect the rather steep change occurring between the morning and the afternoon. For this reason, we are reluctant to consider the best-fitting model as proof that there are two distinct peaks and/or troughs during the 24 h day.

Our analytical approach treats sleep as a confounder to be eliminated, in order to best estimate diurnal variation. The effect of sleep on circulating IL-6 is an interesting question in its own right, but is better addressed by experimental sleep deprivation studies, which were not investigated here. A recent meta-analysis of the effect of sleep deprivation on IL-6 included 12 studies and found no significant effect [20].

One previous meta-analysis [9] (reported again in [10] and [11]) has investigated diurnal variation of circulating IL-6. As discussed in the introduction, this earlier meta-analysis aimed primarily to describe diurnal variation in patients with rheumatoid arthritis, and no systematic method to find data from healthy participants was described. Compared to this earlier meta-analysis, we have used a more systematic approach and we include more data ($k = 56$ datasets compared to $k = 11$). The present findings contrast markedly with those of the earlier meta-analysis, as that study located a morning peak at approx. 06:00 in healthy controls (Fig 1) as well as in patients with rheumatoid arthritis, while we report a morning trough at about 08:00-09:00 in healthy humans. This raises the hypothesis that diurnal variation is different between healthy humans and patients with rheumatoid arthritis, possibly correlating in patients to the diurnal time course of symptoms such as joint stiffness and pain, which tend to be worse in the morning.

The shape of the diurnal curve estimated in this meta-analysis is not suggestive of a mechanism where the immune system secretes more IL-6 into the blood at night in order to promote sleepiness. While the present results do not disprove this putative mechanism, the lack of a night-time peak, compared to the afternoon, suggests that other regulatory mechanisms are dominant.

The estimated morning trough is rather close after the time of day when cortisol levels peak, and also close to the diurnal trough of monocyte and lymphocyte concentrations in peripheral blood [95–97]. The data investigated here cannot speak directly to the relationship between diurnal variation of IL-6 to cortisol and white blood cell concentrations, and further studies will be required to elucidate whether any direct links exist.

The diurnal variation estimated here is large enough to pose a risk of confounding if sampling is performed without regard to time of day, and we therefore recommend that time of day should be taken into consideration in studies recording IL-6 in plasma or serum from healthy humans. Further research is required to determine conclusively whether other cytokines also show diurnal variation. As far as we are aware, IL-6 is the only cytokine to date to be subject to a meta-analysis of diurnal variation.
Author Contributions

Conceived of the study: GN. Designed the study: GN, MI. Collected data: GN. Analysed data: GN, MI. Interpreted results: GN, TÅ, ML, JA, MI. Drafted the manuscript: GN. All authors read and approved the final version of the manuscript.

Competing interests

The authors have no competing interests to declare.

References


20. Irwin MR, Olmstead R, Carroll JE. Sleep Disturbance, Sleep Duration, and Inflammation: A Systematic Review and Meta-Analysis of Cohort Studies and Experimental Sleep Deprivation. Biological Psychiatry;doi:10.1016/j.biopsych.2015.05.014.


Table 1. Characteristics of included studies. Where studies reported several datasets, these are specified separately.

<table>
<thead>
<tr>
<th>1st Author</th>
<th>Year</th>
<th>Ref.</th>
<th>n_{subjects}</th>
<th>n_{timepoints}</th>
<th>Regression weight, %</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sothern</td>
<td>1995</td>
<td>[13]</td>
<td>11</td>
<td>8</td>
<td>0.99</td>
<td>a</td>
</tr>
<tr>
<td>Späth-Schalbe</td>
<td>1998</td>
<td>[32]</td>
<td>16</td>
<td>4</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>Dugué</td>
<td>1998</td>
<td>[33]</td>
<td>22</td>
<td>3</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>1999</td>
<td>[19]</td>
<td>8</td>
<td>24</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>Úndar</td>
<td>1999</td>
<td>[28]</td>
<td>10</td>
<td>7</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Späth-Schalbe</td>
<td>2000</td>
<td>[34]</td>
<td>18</td>
<td>3</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Redwine</td>
<td>2000</td>
<td>[35]</td>
<td>31</td>
<td>15</td>
<td>3.80</td>
<td>b</td>
</tr>
<tr>
<td>Haack</td>
<td>2000</td>
<td>[36]</td>
<td>20</td>
<td>9</td>
<td>1.90</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2000</td>
<td>[37]</td>
<td>12; 11</td>
<td>2; 2</td>
<td>0.54; 0.49</td>
<td>c</td>
</tr>
<tr>
<td>Baker</td>
<td>2001</td>
<td>[38]</td>
<td>8</td>
<td>6</td>
<td>0.62</td>
<td></td>
</tr>
<tr>
<td>Haack</td>
<td>2001</td>
<td>[39]</td>
<td>10</td>
<td>10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2002</td>
<td>[26]</td>
<td>11</td>
<td>24</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2003</td>
<td>[25]</td>
<td>15; 13</td>
<td>48; 48</td>
<td>3.29; 2.85</td>
<td></td>
</tr>
<tr>
<td>Alberti</td>
<td>2003</td>
<td>[40]</td>
<td>20</td>
<td>2</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td>Irwin</td>
<td>2004</td>
<td>[41]</td>
<td>15</td>
<td>4</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2004</td>
<td>[42]</td>
<td>25</td>
<td>48</td>
<td>5.49</td>
<td></td>
</tr>
<tr>
<td>Szezudlik</td>
<td>2004</td>
<td>[43]</td>
<td>17</td>
<td>4</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td>Alesci</td>
<td>2005</td>
<td>[44]</td>
<td>23</td>
<td>9</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>Mehra</td>
<td>2006</td>
<td>[45]</td>
<td>150; 23; 35; 30</td>
<td>2; 2; 2; 2; 2</td>
<td>6.72; 1.03; 0.54; 1.56; 1.34</td>
<td>d</td>
</tr>
<tr>
<td>Burgos</td>
<td>2006</td>
<td>[46]</td>
<td>11</td>
<td>8</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2007</td>
<td>[47]</td>
<td>45</td>
<td>20; 20</td>
<td>4.39; 4.39</td>
<td></td>
</tr>
<tr>
<td>Frey</td>
<td>2007</td>
<td>[48]</td>
<td>19</td>
<td>15</td>
<td>2.33</td>
<td>b</td>
</tr>
<tr>
<td>Lindahl</td>
<td>2007</td>
<td>[49]</td>
<td>14</td>
<td>4</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Miles</td>
<td>2008</td>
<td>[50]</td>
<td>51</td>
<td>5</td>
<td>3.61</td>
<td></td>
</tr>
<tr>
<td>Vgontzas</td>
<td>2008</td>
<td>[51]</td>
<td>48</td>
<td>15; 13</td>
<td>48; 48</td>
<td>3.29; 2.85</td>
</tr>
<tr>
<td>Knudsen</td>
<td>2008</td>
<td>[52]</td>
<td>30</td>
<td>15</td>
<td>1.14</td>
<td>a, f</td>
</tr>
<tr>
<td>Benedict</td>
<td>2009</td>
<td>[53]</td>
<td>19</td>
<td>17</td>
<td>2.09</td>
<td></td>
</tr>
<tr>
<td>Peeling</td>
<td>2009</td>
<td>[54]</td>
<td>8</td>
<td>5</td>
<td>0.57</td>
<td></td>
</tr>
<tr>
<td>Phillips</td>
<td>2010</td>
<td>[55]</td>
<td>7</td>
<td>3; 3</td>
<td>0.38; 0.38</td>
<td>g</td>
</tr>
<tr>
<td>Rief</td>
<td>2010</td>
<td>[56]</td>
<td>52</td>
<td>60; 52</td>
<td>2; 2</td>
<td>2.69; 2.33</td>
</tr>
<tr>
<td>Gill</td>
<td>2010</td>
<td>[57]</td>
<td>14</td>
<td>13</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>Chenmaoui</td>
<td>2011</td>
<td>[58]</td>
<td>12</td>
<td>6</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Pledge</td>
<td>2011</td>
<td>[59]</td>
<td>6</td>
<td>4</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Crispim</td>
<td>2012</td>
<td>[60]</td>
<td>6; 7; 9</td>
<td>7; 7</td>
<td>0.50; 0.59; 0.75</td>
<td></td>
</tr>
<tr>
<td>Voderholzer</td>
<td>2012</td>
<td>[61]</td>
<td>57</td>
<td>16</td>
<td>1.24</td>
<td></td>
</tr>
<tr>
<td>Grigoleit</td>
<td>2012</td>
<td>[62]</td>
<td>58</td>
<td>10</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>Abedelmalek</td>
<td>2013</td>
<td>[63]</td>
<td>59</td>
<td>12</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>Pejovic</td>
<td>2013</td>
<td>[64]</td>
<td>60</td>
<td>30</td>
<td>4.65</td>
<td></td>
</tr>
<tr>
<td>Lekander</td>
<td>2013</td>
<td>[65]</td>
<td>61</td>
<td>9</td>
<td>1.30</td>
<td>a</td>
</tr>
<tr>
<td>Agorastos</td>
<td>2014</td>
<td>[66]</td>
<td>27</td>
<td>11</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>Kritikou</td>
<td>2014</td>
<td>[67]</td>
<td>62</td>
<td>18; 21</td>
<td>24; 24</td>
<td>2.79; 3.26</td>
</tr>
<tr>
<td>Karshikoff</td>
<td>2015</td>
<td>[68]</td>
<td>14</td>
<td>21</td>
<td>4</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Total: 1100

n_{subjects} and n_{timepoints} indicate the number of participants and timepoints, respectively, for each study. The notes section provides additional information about each study, including whether individual participant data were available, if some data were re-coded using mean chronological time as a best approximation, if averaged over 3 consecutive days, if 358 of 385 participants were included in analyses of IL-6, if final n for each sub-group was not given, if each dataset was said to have 50% of the total participants (n = 41), and if the same 7 participants were included twice with a 10-week interval, yielding two different data sets. Further details are provided in the notes section.
<table>
<thead>
<tr>
<th>1st author</th>
<th>Year</th>
<th>Ref.</th>
<th>n\textsubscript{subjects}</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lemmer</td>
<td>1992</td>
<td>64</td>
<td>12</td>
<td>Data could not be estimated (below detection limit)</td>
</tr>
<tr>
<td>Gudewill</td>
<td>1992</td>
<td>7</td>
<td>12</td>
<td>Data could not be estimated (given as counts)</td>
</tr>
<tr>
<td>Pollmächer</td>
<td>1993</td>
<td>65</td>
<td>15</td>
<td>Data could not be estimated (not shown; were “close to assay detection limit”)</td>
</tr>
<tr>
<td>Bauer</td>
<td>1994</td>
<td>8</td>
<td>5</td>
<td>Data could not be estimated (given as arbitrary units)</td>
</tr>
<tr>
<td>Dinges</td>
<td>1994</td>
<td>66</td>
<td>20</td>
<td>Data could not be estimated (not shown)</td>
</tr>
<tr>
<td>Arvidson</td>
<td>1994</td>
<td>65</td>
<td>10</td>
<td>Data could not be estimated (below detection limit)</td>
</tr>
<tr>
<td>Seiler</td>
<td>1994</td>
<td>15</td>
<td>6</td>
<td>Data could not be estimated (clock time not given, and too low resolution)</td>
</tr>
<tr>
<td>Seiler</td>
<td>1995</td>
<td>68</td>
<td>6</td>
<td>Data same as in [15]</td>
</tr>
<tr>
<td>Sothern</td>
<td>1995</td>
<td>23</td>
<td>10</td>
<td>Data same as in [13]</td>
</tr>
<tr>
<td>Pollmächer</td>
<td>1996</td>
<td>69</td>
<td>20</td>
<td>Data could not be estimated (given as difference between treatments)</td>
</tr>
<tr>
<td>Korth</td>
<td>1996</td>
<td>70</td>
<td>20</td>
<td>Data could not be estimated (too low resolution)</td>
</tr>
<tr>
<td>Crofford</td>
<td>1997</td>
<td>71</td>
<td>5</td>
<td>Data could not be estimated (largely below detection limit)</td>
</tr>
<tr>
<td>Gudmundsson</td>
<td>1997</td>
<td>72</td>
<td>1</td>
<td>Very high levels</td>
</tr>
<tr>
<td>Born</td>
<td>1997</td>
<td>73</td>
<td>10</td>
<td>Very high levels</td>
</tr>
<tr>
<td>Lissoni</td>
<td>1998</td>
<td>74</td>
<td>10</td>
<td>Data could not be estimated (largely below detection limit)</td>
</tr>
<tr>
<td>Bornstein</td>
<td>1998</td>
<td>75</td>
<td>9</td>
<td>Data could not be estimated (not shown)</td>
</tr>
<tr>
<td>Hermann</td>
<td>1998</td>
<td>76</td>
<td>10</td>
<td>Data could not be estimated (too low resolution)</td>
</tr>
<tr>
<td>Kanabrocki</td>
<td>1999</td>
<td>21</td>
<td>11</td>
<td>Data same as in [13]</td>
</tr>
<tr>
<td>Genesca</td>
<td>2000</td>
<td>74</td>
<td>8</td>
<td>Data could not be estimated (largely below detection limit)</td>
</tr>
<tr>
<td>Mastorakos</td>
<td>2000</td>
<td>75</td>
<td>5</td>
<td>Data same as in [71]</td>
</tr>
<tr>
<td>Mullington</td>
<td>2000</td>
<td>79</td>
<td>19</td>
<td>Data could not be estimated (given as change between conditions), also possibly same as in [70]</td>
</tr>
<tr>
<td>Johansson</td>
<td>2000</td>
<td>80</td>
<td>18</td>
<td>Data could not be estimated (largely below detection limit)</td>
</tr>
<tr>
<td>Shaw</td>
<td>2001</td>
<td>81</td>
<td>10</td>
<td>Data could not be estimated (largely below detection limit)</td>
</tr>
<tr>
<td>Lange</td>
<td>2002</td>
<td>82</td>
<td>18</td>
<td>Data could not be estimated (given as z-transformed change between time points)</td>
</tr>
<tr>
<td>Domínguez-Rodríguez</td>
<td>2003</td>
<td>83</td>
<td>40</td>
<td>Very high levels</td>
</tr>
<tr>
<td>Domínguez-Rodríguez</td>
<td>2004</td>
<td>84</td>
<td>60</td>
<td>Very high levels and possibly overlapping with [83]</td>
</tr>
<tr>
<td>Haack</td>
<td>2007</td>
<td>85</td>
<td>18</td>
<td>Data could not be estimated (given as change between conditions)</td>
</tr>
<tr>
<td>Eisenberger</td>
<td>2009</td>
<td>86</td>
<td>16</td>
<td>Data could not be estimated (too low resolution)</td>
</tr>
<tr>
<td>Eisenberger</td>
<td>2010</td>
<td>87</td>
<td>16</td>
<td>Data same as in [86]</td>
</tr>
<tr>
<td>Haimovich</td>
<td>2010</td>
<td>88</td>
<td>2</td>
<td>Data could not be estimated (given as change between conditions)</td>
</tr>
<tr>
<td>Sauvet</td>
<td>2010</td>
<td>89</td>
<td>12</td>
<td>Data same as in [54]</td>
</tr>
<tr>
<td>Grigoleit</td>
<td>2011</td>
<td>90</td>
<td>34</td>
<td>Data could not be estimated (control condition not shown)</td>
</tr>
<tr>
<td>Miles</td>
<td>2012</td>
<td>91</td>
<td>30</td>
<td>Data same as in [31]</td>
</tr>
<tr>
<td>Schrepf</td>
<td>2014</td>
<td>92</td>
<td>28</td>
<td>Data could not be estimated (time of sampling not shown)</td>
</tr>
<tr>
<td>Wegner</td>
<td>2014</td>
<td>93</td>
<td>18</td>
<td>Data could not be estimated (too low resolution)</td>
</tr>
<tr>
<td>Scott</td>
<td>2015</td>
<td>94</td>
<td>14</td>
<td>Data could not be estimated (clock time not given)</td>
</tr>
</tbody>
</table>

The total number of participants does not count twice those participants included in duplicate reports.
Figure 1

Note to typesetter: The numbers in this figure need to match the reference list. Suggested workflow is that you typeset the manuscript without regard to this figure, send us the proofs, and we then make a final version of the figure with numbers in correspondence to the reference list as it will then be ordered.
Figure 2

Identified in database search: $k = 507$

Screened by titles and abstracts: $k = 507$

Ineligible: $k = 440$

Eligible: $k = 66$

- Data could not be estimated ($k = 25$)
- Duplicate publication ($k = 7$)
- Very high levels ($k = 4$)

Entered in quantitative synthesis: $k = 43$

Hand-searching of reference lists and conference proceedings etc: $k = 13$
Figure 3
Figure 4
Figure 6
Figure 7