Optimizing the Assessment of Parental Burnout: A Multi-Informant and Multi-Method Approach to Determine Cut-offs for the Parental Burnout Inventory (PBI) and the Parental Burnout Assessment (PBA)

Maria Elena Brianda[1,2]*, Moïra Mikolajczak[1], Michel Bader[3], Sandra Bon[4], Alexandra Déprez[5], Nicolas Favez[6], Liliane Holstein[7], Sarah Le Vigouroux[8], Astrid Lebert-Charron[5], Raquel Sánchez-Rodríguez[9], Natalène Séjourné[9], Jaqueline Wendland[5], Isabelle Roskam[1]

[1]Psychological Sciences Research Institute, UCLouvain, (Belgium)

[2]Research Unit for a life-Course perspective on Health and Education (RUCHE), Department of Psychology, University of Liège (Belgium)

[3]University Service of Child and Adolescent Psychiatry, University Hospital of Lausanne (Switzerland)

[4]Swiss Association for Cognitive Psychotherapy, Geneva (Switzerland)

[5]Laboratory of Psychopathology and Health Processes, Paris Descartes University (France)

[6]Faculty of Psychology and Educational Sciences, University of Geneva (Switzerland)

[7]Psychoanalyst in independent practice, Boulogne (France)

[8]UNIV.NIMES, EA 7352, CHROME, Team APSY^{-v} (France)

[9]Center for Studies and Research in Psychopathology and Health Psychology, University of Toulouse (France)

*Corresponding author: Maria Elena Brianda, University of Liège, 2 Place des Orateurs (Quartier Agora), 4000 Liège, Belgium, mariaelena.brianda@uliege.be, ORCID 0000-0001-5566-3184

Abstract

Parental Burnout (PB) is a chronic stress-related condition resulting from long-lasting exposure to overwhelming parenting stress. Previous studies showing the seriousness of this condition stressed the urgent need to provide researchers and practitioners with effective assessment tools. Validated PB measures are the Parental Burnout Inventory (PBI) and the Parental Burnout Assessment (PBA). The good psychometric properties of these instruments have been replicated across different samples and countries, but thresholds for identifying impairing PB levels (i.e., cut-off scores) have not yet been established. The present study aims to fill this gap by adopting a multi-informant and multi-method approach to a sample of 192 burned-out and control parents. PBI and PBA cut-offs were derived from the combination of several PB indicators, based on a preregistered analysis strategy. Results identified a score of 74.6 (95% CI [69.48 – 79.68]) for the PBI and 86.3 (95% CI [79.49 – 93.03]) for the PBA as indicators of the most severe PB levels.

Keywords: burnout, parent, assessment, multi-method, cut-off, diagnosis

Introduction

Parental Burnout (PB) is a clinically significant condition resulting from long-term exposure to overwhelming parenting stress (Mikolajczak & Roskam, 2018; Roskam et al., 2017). It involves emotional exhaustion in one's parental role, an emotional distancing toward children, the loss of pleasure of being with them, and the feeling of not being a good parent anymore (Roskam et al., 2018). The severity of PB consequences for parents (e.g., increased risk of suicidal ideation, dysregulation of the hypothalamic-pituitary-adrenal axis) and children (e.g., increased parental neglect and violence) attests to the seriousness of this condition (Brianda, Roskam, & Mikolajczak, 2020; Mikolajczak et al., 2019). These consequences call for both efficient treatments for burned-out parents and assessment tools to measure PB symptoms and indicate their severity. While research has begun to address the need for efficient treatments (Brianda, Roskam, Gross, et al., 2020), the need for assessment tools has been only partially met. Valid instruments to assess PB have been developed (Roskam et al., 2017, 2018) and translated (Aunola et al., 2020; Baldisserotto et al., 2018; Kawamoto et al., 2018), but the absence of clinical cut-offs renders these instruments of limited use to practitioners. The current paper aims to overcome this weakness and provide clinicians with validated cut-off scores on the two most widely used PB measures. These cut-off scores will not only be useful for practitioners but will also provide a useful reference for researchers (for future epidemiological studies and/or to facilitate the interpretation of outcomes in clinical trials).

The two validated measures for the assessment of PB symptoms are the Parental Burnout Inventory (Roskam et al., 2017) and the Parental Burnout Assessment (Roskam et al., 2018). The Parental Burnout Inventory (PBI) comes from an adaptation to the parenting context of the gold standard instrument for assessing job burnout: the Maslach Burnout Inventory (Maslach et al.,

1986). The validity of the PBI and its tridimensional structure (i.e., emotional exhaustion, emotional distancing, and loss of personal accomplishment) has been replicated across different samples and contexts (e.g., samples of mothers and fathers, see Roskam & Mikolajczak, 2020; French-speaking and English-speaking parents, see Roskam et al., 2017 and Roskam & Mikolajczak, 2020; Japanese parents, see Kawamoto et al., 2018; Dutch parents, see Van Bakel et al., 2018). The use of the PBI was recommended in studies aiming to compare burnout in two contexts, i.e., work and family (Roskam et al., 2018). The similar structure of the PBI and the Maslach Burnout Inventory may be helpful to study, for instance, the relevance of the context in which burnout symptoms occur, common and distinct causes and consequences of burnout in the two contexts, and whether burnout remains limited to one context or whether it spreads to multiple spheres of life (Mikolajczak et al., 2020). The Parental Burnout Assessment (PBA) was designed using an inductive method solely based on burned-out parents' testimonies. Its four-dimensional conceptualization of PB (i.e., emotional exhaustion in one's parental role, emotional distancing from one's children, sense of being fed-up with parenting, and contrast with the previous parental self) constitutes so far, the best representation of PB. It has been validated in several languages (e.g., Arabic, Chinese, English, Farsi, Finnish, Japanese, Polish, Portuguese, Romanian, and Spanish[1]) and was chosen as the reference measure in the International Investigation of Parental Burnout (Roskam et al., 2021), an extensive survey on PB intercultural variation involving more than 40 countries across the world (https://www.burnoutparental.com/international-consortium).

Both the PBI and PBA have shown good psychometric properties and good convergent validity (Roskam et al., 2017, 2018; Roskam & Mikolajczak, 2020), but thresholds for identifying parents suffering from impairing PB levels are still missing. In the framework of PBI

and PBA validation studies, several authors attempted to estimate PB prevalence in their study sample (Kawamoto et al., 2018; Roskam et al., 2018, 2017; Van Bakel et al., 2018). Different approaches have been used to identify burned-out parents, i.e., a "theoretical" approach based on the response scale such as displaying at least 66.6% of the PB symptoms every day (Kawamoto et al., 2018; Roskam et al., 2018, 2017; Van Bakel et al., 2018), or a statistical approach corresponding to 1.5 standard deviations above the group mean (Kawamoto et al., 2018; Roskam et al., 2017; Van Bakel et al., 2018), or even an approach based on cut-offs provided for job burnout (Roskam et al., 2017; Van Bakel et al., 2018). As the authors themselves discussed, none of the three methods above appears fully satisfying. First, none of these cut-off methods are based on objective external criteria. Second, they lead to considerably variable prevalence rates of burned-out parents in the same sample (depending on the criteria adopted, PB point prevalence can range from 0.2 to 17.3%).

Another possibility to meet the need to establish PB cut-offs would be to compare parents considered to be suffering from PB vs. control parents to examine PBI and PBA's ability to discriminate between the two categories and derive the most accurate cut-offs. Nevertheless, the absence of internationally recognized validated criteria for the classification of PB scores makes it hard to make decisions on the presence or absence of the syndrome. Another option would also be to compare parents who ask for PB treatment vs. control parents. However, using the criterion of seeking treatment alone risks being misleading. The nature of the PB experience is indeed highly subjective and parents may ask for help even with low PB levels (Brianda, Roskam, & Mikolajczak, 2020). Such a criterion should thus be complemented by external and objective criteria to compensate for the subjectivity limit.

Based on the foregoing, the most appropriate way to establish PB clinical cut-offs is to

rely on a bundle of indicators derived from different informants and various methods, including both subjective and objective external criteria, which provide different points of view on the presence of PB. To this purpose, we will employ a rigorous strategy based on a multi-informant and multi-method assessment that includes: (i) the views of parents and external judges alike, and (ii) a combination of self-reported questionnaires, free speech samples, and a biological marker of chronic stress (the hair cortisol concentration). Both parents asking for clinical treatment for PB and control parents have been included in the data collection. Method and analyses performed in the current study were preregistered on the Open Science Framework on June 13th, 2020, as a preregistration of secondary data analysis (see the Statistical analyses section for further details). The preregistration can be found at https://osf.io/ujfb3.

Method

Participants

For the purpose of the current study, we combined data collected in two subsamples from two previous studies on PB treatment (Brianda, Roskam, Gross, et al., 2020) and PB biological correlates (Brianda, Roskam, & Mikolajczak, 2020). The first subsample (henceforth "subsample 1"; $n = 130$) consisted of parents voluntarily enrolled in group treatment for PB, and the other (henceforth "subsample 2"; $n = 62$) consisted of control parents. The total sample consisted of 192 parents (92.7% mothers) from the French-speaking part of Belgium. The majority were aged between 35 and 44 (53.4%), had two or three children (71.5%), and were in a couple (87%). Most of them (80.2%) were employed, had a bachelor's or master's degree (69.5%), and had a household net monthly income between €2500 and €5500 ($2800 and $6160; 68.9%). The sociodemographic characteristics of the sample are presented in more detail in the Supplemental materials (see Table S1). The subsamples of the two previous studies were statistically

equivalent with respect to sociodemographic characteristics, except for work status. Almost all unemployed parents were part of the sample of the PB treatment study. The effect size of the difference, however, was small (φ=0.29).

Procedure

Inclusion criteria for subsample 1 were having at least one child still living at home and applying for treatment specifically aimed at reducing PB (recruitment and data collection are fully detailed in Brianda et al., 2020, Supplemental materials). Data used in the current study were collected before the beginning of the treatment.

Subsample 2 consisted of control parents willing to participate in a study on the "Estimation of hair cortisol levels in parents" (recruitment and data collection are fully detailed in Brianda, Roskam, & Mikolajczak, 2020). Inclusion criteria were having at least one child still living at home and having hair at least 3 cm long (a necessary pre-condition for the hair cortisol analysis; see the Measures section below).

We collected self-reported measures (provided by participants), clinical judgments (completed by external judges based on a five-minute speech provided by participants on their parenting experience), and a biological measure of stress (the analysis of cortisol levels contained in participants' hair) in both subsamples. To ensure data confidentiality, all participants were identified by anonymous codes. A consent form informed parents about their right to withdraw at any time and/or not provide one or another measure (e.g., the speech sample or the hair sample). For parents of subsample 1 (i.e., those enrolled in the treatment), the consent form made it clear that drop-out from the study would in no way compromise their participation in the treatment.

As a first step, we invited all eligible participants to complete a self-reported assessment of PB online (via a link sent by the researcher). Overall, 192 eligible parents agreed to fill in the protocol. Of these, six participants did not answer the PBI questions, and eight participants did not report the PB level and category. Self-reported measures are fully described in the Measures section below.

After completing the self-reported measures, participants were invited to audiotape a five-minute free speech on their parenting experience. Participants could tape the five-minute speech either at home or on a voice recorder provided by the researcher during the meeting for the collection of the hair sample (in either case, parents of subsample 1 recorded the five-minute speech before the first session of treatment). We made every effort to ensure that recording conditions were similar across different settings. In both cases, we asked the parents to record themselves alone, in a quiet place (in the lab, the parent was left alone in a quiet room where they could record in total privacy). Participants who taped the speech at home were asked to follow the instructions provided at the end of the online questionnaire; participants who taped the speech in the lab received the exact same instructions written on a sheet of paper. Participants received the following instructions: “We're asking you to talk for five minutes about your experience and your feelings as a parent. You can say spontaneously everything that comes to your mind when you think about your parental role”. The duration of the speech (5 minutes) has been chosen based on extensive research showing that (1) a five-minute speech sample provides enough material to allow judgment accuracy (Magaña et al., 1986) and that (2) judgments made relatively quickly based on thin slices of recording can actually be more efficient than judgments pondered on extensive material (see Ambady, 2010, for review). However, although the instructions called for a five-minute speech, in all cases (both at home and in the lab) the

participants freely chose the actual duration (this implies that in the lab, the researcher did not interrupt participants or force them to fill the five minutes). Of the entire sample, 115 parents accepted to provide the five-minute speech. All the audio-taped files have been fully transcribed by a professional data entry company based in another country (London, UK) to avoid the risk that parents could be identified.

We then asked a pool of external judges to assess participants' PB based on listening to the speech samples. We recruited eight judges (i.e., seven psychologists and one psychiatrist) who met at least one of the following criteria: (i) having a master's degree in Psychology or a Psychiatry degree as a minimum level of education, and being a clinician with at least five years of clinical experience with parents (four judges met these criteria), or (ii) having a master's degree in Psychology and being an academic expert in the field of PB and/or parental stress, with at least three years of research experience and at least two publications as the first author on this domain (four judges met these criteria). We recruited French-speaking judges outside Belgium, i.e., in France, Switzerland, and Luxembourg, to limit the chance they could identify parents based on their voice and personal details provided in the free speech. This enabled us to ensure participants that parents' data would remain confidential. The principal investigators of this study (i.e., the first, second, and last author) were not included among the judges to avoid any bias.

We randomly assigned the 115 five-minute free speeches to the eight judges. Each speech was assigned to two different judges for interrater reliability purposes (see the Statistical analyses section below). Three pairs of judges received 29 five-minute speeches, and the fourth pair received 28 speeches to analyze, with for all of them 50% coming from parents asking for treatment and 50% coming from control parents (in completely random order). Judges were blind

to the subsample from which the parents came, and they did not know the percentage of burned-out and control parents in their batch. Each pair was made up of a clinical judge and an academic judge so that each speech sample would be evaluated with the two forms of expertise. We provided judges with both the audio-taped files and their transcripts. We invited them to carefully listen to them and to answer a short questionnaire about participants' PB right after (further details are provided in the Measures section below). Judges did not know their pair and realized the evaluations independently from each other.

Finally, we measured participants' hair cortisol concentration (HCC) because burnout is a chronic stress condition and hair cortisol is a biomarker of chronic stress (Stalder & Kirschbaum, 2012). Hair sampling for the HCC assessment took place on the day of the first session of the treatment for parents of subsample 1 or during an appointment expressly set up with the researcher for the parents of subsample 2. Of the entire sample, 184 parents accepted to provide the hair sample. Each participant provided a strand of approximately 150 hairs (i.e., a strand with a diameter of at least 3 mm or 1/8 inch, which corresponds to the diameter of half of a pencil), collected from the posterior vertex of their head. We cut hair samples proximal to the scalp, wrapped them in aluminum foil, and stored them in an envelope. Hair samples were then sent to a specialized laboratory at the University of Granada for analysis (more details about HCC analysis are provided in the Measures section below).

Measures

Sociodemographic and health-related characteristics. Participants provided the following sociodemographic information: gender, age, race/ethnicity, number of children, marital status, educational level, work status, and net monthly household income. We also asked

participants about medication intake (and in particular oral cortisone) for the examination of hair cortisol levels.

Self-reported measures. Self-reported questionnaires aimed at gathering participants' perceptions of experienced PB. Participants first indicated PB category and level and then completed the PBI and PBA questionnaires. The category, level, and questionnaire scores of PB provide three different kinds of information. Although we expect that in most cases, these will point in the same direction (e.g., parents who rate the PB category as "moderate" are likely to choose a relatively low score on a scale of zero to ten and will also score low on questionnaires), this is not always true. Some parents may score high on PB questionnaires and yet indicate their PB as "moderate", in cases where their symptoms although frequent, are not perceived as severe. In other cases, parents may rate their PB level using a high score on a scale from zero to ten while obtaining a moderate score on the questionnaires, perhaps due to a phenomenon of social desirability when confronted with the harsh reality expressed in the items. We believe that each of these indicators gives us valuable knowledge about the parent's suffering. In the absence of other validated questionnaires for the self-reported assessment of PB, we, therefore, decided to consider PB category and level in addition to PB questionnaires to collect participants' subjective perceptions of the severity and magnitude of their condition.

PB category. We asked participants which one of the following categories corresponds best to their actual state ("If you were to place the severity of your parental exhaustion in one of the categories below, you would say that you are..."): *not in PB* (0), *minor PB* (1), *moderate PB* (2), or *severe PB* (3).

PB level. We asked participants to indicate their degree of PB ("Could you visually indicate your degree of parental burnout on the gauge below?") on a scale from 0 (*not in PB at*

all) to 10 (*extreme PB*). They could report their level of PB by placing the pointer of a graduated dial (Figure A1 in the Appendix shows the gauge displayed in the online questionnaire).

PB scores. We invited participants to complete the two existing validated questionnaires for the assessment of PB symptoms: the Parental Burnout Inventory – PBI (Roskam et al., 2017) and the Parental Burnout Assessment - PBA (Roskam et al., 2018). The PBI is a 22-item questionnaire deductively derived from the Maslach Burnout Inventory (Maslach et al., 1986) and adapted to the context of parenting. The PBA is a 23-item questionnaire created through an inductive method based solely on the testimonies of burned-out parents. In both questionnaires, parents are invited to rate the presence of each PB symptom/item on a seven-point frequency scale: *never* (0), *a few times a year or less* (1), *once a month or less* (2), *a few times a month* (3), *once a week* (4), *a few times a week* (5), and *every day* (6). PBI total score can therefore range from 0 to 132, while PBA from 0 to 138. Higher scores indicate higher levels of PB. Cronbach's alphas in the current sample were 0.97; 95% CI [0.96, 0.98] for PBI and 0.98; 95% CI [0.98, 0.99] for PBA.

Clinical judgments based on the five-minute speech. We asked judges to estimate PB's presence, category, and level. In the absence of validated diagnostic tools for the clinical judgment of PB, we employed the same measures as for the self-reported assessment. We held scales and response labels constant across informants (i.e., parents and judges) to minimize the amount of discrepancy due to measurement (De Los Reyes et al., 2013).

PB category. We asked external judges which one of the following categories they would choose if they had to rank the category of participants' PB ("If you were to place the severity of the participant's parental exhaustion in one of the categories below, you would say that s/he is..."): *not in PB* (0), *minor PB* (1), *moderate PB* (2), or *severe PB* (3).

PB level. We asked external judges to indicate participants' degree of PB ("Could you visually indicate the participant's degree of parental burnout on the gauge below?") on a scale from 0 (*not in PB at all*) to 10 (*extreme PB*). We used the same gauge as on the self-reported protocol (see Figure A1 in the Appendix).

Hair Cortisol Concentration (HCC). HCC in hair samples was analyzed using the Salivary ELISA Cortisol kit (Russell et al., 2015). The laboratory analyzed only the 3 cm most proximal to the scalp, which provide a measure of cortisol accumulation over the three previous months (Staufenbiel et al., 2013). The procedure for the analysis of HCC is fully detailed in Caparros-Gonzalez et al. (2017).

Statistical Analyses

We included in the analyses all the participants whose available data included the sociodemographic and health-related characteristics, the PBI or the PBA, and at least one of the following indicators: self-reported PB category, self-reported PB level, clinical judgments about PB category, clinical judgments about PB level, or HCC. We performed binary logistic regressions to check whether there was any significant predictor of data missingness. We investigated the effect of gender, age, race/ethnicity, marital status, number of children, educational level, work status, and net monthly income on the binary variable indicating missing data for each variable under study. Results showed that none of the possible predictors considered explained the likelihood of having missing data, suggesting that data were missing at random. Participants who used oral cortisone have been excluded from the analyses considering the HCC variable.

The first step consisted in distinguishing parents suffering from PB from others on each indicator of PB (except for PBI and PBA, for which we were seeking to derive cut-off scores): the fact of seeking treatment for PB, the self-reported PB category, the self-reported PB level, the clinical judgments about PB category, the clinical judgments about PB level, and the HCC. To do so, we computed six dichotomous variables (i.e., one for each indicator) that classified parents into two categories: parents most likely suffering from PB (i.e., cases, value "1"), vs. parents most likely not suffering from PB (i.e., controls, value "0"). The dichotomous criteria below were considered as possible indicators that the parent was most likely suffering from PB (i.e., cases):

- Indicator *a*: the participant voluntarily enrolled in a treatment aimed at reducing PB symptoms;
- Indicator *b*: the participant categorized their PB as "moderate" or "severe" [2];
- Indicator *c*: the participant reported experiencing a level of PB of at least seven on a scale from zero (*not in PB at all*) to 10 (*extreme PB*)[3];
- Indicator *d*: at least one judge categorized the participant's PB as "moderate" or "severe"[2];
- Indicator *e*: the mean score computed between the PB levels reported by the two judges (i.e., the mean between the PB level indicated by the first judge and the PB level indicated by the second judge) was equal to or greater than seven[3];
- Indicator *f*: HCC found in the participant's hair sample was greater than or equal to 75.9 pg/mg hair[4].

As regards indicators *d* and *e* (those based on the clinical judgments), given that every participant was assessed twice by two independent judges, we first checked for interrater

reliability based on the judges' assessment of the PB category. We considered clinical judgments as reliable in two cases only: when the two judges attributed to the parent the same category of PB (e.g., both considered the parent to be "not in PB", or in a "severe PB"; 72 clinical judgments out of 115 met this condition), or when their judgment differed by only one category (e.g., one judge said "not in PB" while the other said "minor PB", or one said "moderate PB" and the other said "severe PB"; 35 out of 115 clinical judgments met this condition). We, therefore, dismissed from the analyses clinical judgments when a discrepancy of two or three categories was found between judges' evaluations (e.g., one judge said "not in PB" while the other said "moderate PB", or one judge said "minor PB" and the other said "severe PB"). Overall, only 6.9% of the clinical judgments, i.e., eight cases in total, were dismissed. In retained cases, we also observed excellent interrater consistency for the ratings of participants' PB levels (Intra-Class Correlation = 0.93).

In a second step, we used the six classifications into cases vs. controls (i.e., the dichotomous variables *a* → *f*) to derive possible cut-offs for the PBI and PBA, respectively, by using two different analysis strategies: (i) the Receiver-Operating Characteristic (ROC) analysis, and (ii) the analysis of means. (i) The ROC analysis is the most widely used procedure to achieve cut-off scores and assess the diagnostic properties of tests (Pintea & Moldovan, 2009). This procedure allows finding the scores of a test that are associated with the highest sensitivity (i.e., the probability that a test result will be positive when the condition is present, also called true positive rate) and the highest specificity (i.e., the probability that a test result will be negative when the condition is not present, also called true negative rate), according to a valid classification that differentiates subjects with or without the condition. ROC analysis is based on the ROC curve, a graph of sensitivity versus 1 - specificity. In our case, we had two tests (the

PBI and the PBA) and six classifications into the presence/absence of impairing PB levels. We thus performed six ROC analyses for each test (i.e., the PBI and the PBA). We examined the area under the ROC curve (AUC) to measure the overall ability of PBI and PBA to discriminate between PB impairing levels and non-impairing levels with respect to each classification. Following Streiner and Cairney (2007), an AUC between 0.50 and 0.70 indicates low accuracy of the questionnaire, an AUC between 0.70 and 0.90 a moderate accuracy, and an AUC over 0.90 indicates high accuracy. Then, we looked for PBI and PBA scores associated with high sensitivity and high specificity with respect to each classification. We followed two widely used approaches for the identification of cut-off points based on sensitivity and specificity: the Closest-to-(0,1) criterion and the Youden index (Akobeng, 2007; Fluss et al., 2005; Perkins & Schisterman, 2006). The Closest-to-(0,1) criterion allows for identifying the cut-off that most closely approximates the performance of a test that perfectly discriminates between cases and controls. On the graph, the curve of a "perfect" test would consist of a vertical line running from (0.0) to (0.1) that joins with a horizontal line running from (0.1) to (1.1). The cut-off determined with the closest to (0.1) criterion corresponds to the point on the ROC curve closest to the (0.1) point, i.e., the graph's upper right corner. The Youden index is a summary statistic of the ROC curve used to identify the cut-off that maximizes the discriminatory ability of the test when equal weight is given to sensitivity and specificity. On the graph, the Youden index corresponds to the point of the maximum vertical distance between the ROC curve and the chance line, which represents the inability to discriminate between cases and controls. In the event that the two approaches led to different scores, we chose the highest value (i.e., the most conservative one). Indeed, as the cut-off increases, the specificity increases as well, thus improving the detection of parents actually suffering from the most severe PB levels and avoiding a wide number of false

positives (Park et al., 2004). In this way, we obtained six potential cut-offs for PBI and six for PBA. (ii) As a second analysis strategy, we computed the PBI and PBA mean scores of parents who received the value of 1 (i.e., classified as cases) on each indicator. We decided to add a second analysis strategy because the sole use of ROC analyses could be risky in this context given the lack of a robust and validated diagnostic criterion. The analysis of means strategy, on its side, has the advantage of yielding values that represent the real experience of parents suffering from the most impairing PB levels in our sample. Such values may be of great importance to researchers and clinicians in identifying the most compromised parents.

We thus obtained six mean scores for each questionnaire (i.e., six mean PBI scores and six mean PBA scores), which represent the six potential cut-offs for the PBI and PBA, respectively. In a third step, we computed the mean of the six potential cut-offs derived from the two analysis strategies and the associated 95% confidence intervals. We thus obtained two cut-offs for each test (i.e., one average cut-off obtained from the ROC analysis, and one obtained from the analysis of means, for both PBI and PBA).

Of note, we opted for computing the average of the scores derived from the six indicators with the aim of capturing the full extent of available information on participants' PB (Augenstein et al., 2016). We believe that the most relevant score should include in its estimation multiple information (i.e., self-reports, clinical judgments, and biological measures) rather than seeking the most valid (as would be the case using modal values or regression analyses to identify the most predictive scores). In this last case, in fact, we would lose meaningful information coming from the variation among different sources (De Los Reyes et al., 2013). Our strategy stems from the joint discussion of 13 experts (i.e., the authors) and relies on the assumption that integrating various sources of information might counterbalance the biases imputable to single sources

(Alexander et al., 2017). Besides this, we made some measures weigh more than others (see Appendix Figure A2). We assigned, for instance, the greatest weight (i.e., three out of six indicators) to self-reported measures, since we considered that parents were the main experts on their experience as a parent, and thus best placed to assess their PB (Demetriou et al., 2015). We assigned instead the lowest weight (i.e., one out of six indicators) to the biological marker because although providing an objective and bias-free measure of chronic stress, we could not exclude that hair cortisol levels were related to other sources of enduring stress (Semmer et al., 2003).

Lastly, we kept for each test the cut-off derived from the strategy that led to the most conservative values (i.e., the highest). Our goal was indeed to find the most relevant clinical cut-off that allows detecting those parents who are suffering from the most severe PB levels, and not to achieve PB overdetection because of a too-low cut-off. Nevertheless, the less conservative values (i.e., the lowest) that will emerge from the analysis strategies described above have been considered risk indicators of moderate PB severity to indicate parents for whom a more in-depth assessment is needed to sharpen therapeutic decision-making.

We preregistered the full data analyses procedure described above on the Open Science Framework on June 13th, 2020 (https://osf.io/ujfb3) before running the analyses because we did not want the strategy to be derived from or adapted to the results emerging from the data[5]. The following results were obtained from the strict application of the preregistered analysis procedure.

401 # Results

402 ## Identification of Cases and Controls According to the six Indicators

403 Table 1 shows the percentage of parents in the total sample classified as cases (i.e., most
404 likely suffering from impairing PB levels) according to the six indicators described in the
405 Statistical analyses section. The percentages are high, but one should remember that this study
406 aims at setting clinical cut-offs (hence, more than half of the current sample is composed of
407 parents seeking treatment for PB), it is not an epidemiological study on the prevalence of PB in
408 the general population. The classifications into cases and controls based on self-reported
409 measures and clinical judgments were moderate to highly correlated (ϕ = .61 to .91). This was
410 not the case for indicator *f* (i.e., the indicator based on HCC), whose classification was weakly
411 correlated to that of the other indicators (all Phi coefficients of correlation between
412 classifications into cases and controls are presented in Supplemental materials Table S2). 41.7%
413 of parents in the total sample were attributed the same classification (either case or control) by
414 *all* the indicators. Comparisons on sociodemographic variables between groups based on the six
415 classifications are presented in Supplemental materials Table S3.

416 ## Overall PBI and PBA Ability to Discriminate Between Cases and Controls

417 The examination of the AUC within the ROC analyses revealed that both the PBI and the
418 PBA globally showed high accuracy in discriminating between cases (i.e., parents suffering from
419 impairing PB levels) and controls with respect to classifications *a* to *e* (AUCs ranging from 0.88
420 to 0.98; see Table 2). Conversely, as regards the classification based on indicator *f*, both tests
421 showed a low accuracy, with no statistically significant AUC (0.58 and 0.57, respectively).
422 Supplemental materials Figures S1 and S2 provide a graphic representation of the ROC curves
423 obtained for PBI and PBA, respectively, with respect to the six classifications of

presence/absence of impairing PB levels. On each graph, we also marked in red the point on the curve that corresponds to the cut-off score resulting from the ROC analysis for PBI (Figure S1) and PBA (Figure S2).

PBI and PBA Cut-offs Resulting From the ROC Analysis and the Analysis of Means

After calculating the mean of the six potential cut-offs derived from the two analysis strategies and the associated 95% confidence intervals, the most conservative (i.e., the highest) average cut-off values were those that resulted from the analysis of means strategy: 74.58, 95% CI [69.48 – 79.68] for the PBI, and 86.26; 95% CI [79.49 – 93.03] for the PBA (Table 3).

Supplementary Analyses

As a supplement to the preregistered plan, we performed some extra analyses to check whether the unequally distributed variable "work status" affected the study results. To this end, we repeated the entire analysis procedure on a subsample randomly generated from the total sample and paired on the variable "work status." The paired sample consisted of 62 parents voluntarily enrolled in a PB treatment and 62 controls, with the same percentages of employed and unemployed participants in it (thus removing the confounding effect of this variable). The extra analyses on the paired sample led to similar results to those obtained on the total sample for the analysis strategy based on the analysis of means, but not for the ROC analyses. We then went one step further and repeated both analysis strategies on a homogeneous sample consisting only of employed parents. This homogenous sample consisted of 154 parents (94 parents voluntarily enrolled in a PB treatment and 60 controls). Results obtained from the homogenous sample were highly comparable to those obtained from the total sample, and this was true for both analysis strategies (the gap between the cut-offs obtained in the two conditions ranges from 0.17 to 4.50 points). These supplementary results allow us to confirm the relevance of the cut-offs obtained

from the total sample presented above. Full results obtained from the paired and the homogenous sample, including sensitivity and specificity values associated with each cut-off, are provided in Supplemental materials, Tables S4, S5, S6, and S7).

Further additional analyses compared with the preregistered plan were carried out to shed light on the influence of the biological stress measure (HCC) on the results. Indeed, the ROC analysis applied to the classification into cases vs. controls based on the biological indicator suggested that HCC might not be able to discriminate between higher and lower levels of PB symptoms as assessed via the PBI and PBA. We have thus performed extra analyses to check what cut-offs we would have obtained by excluding cortisol data from the analyses. We were reassured to find that these cut-offs were not significantly different from those identified by including all the PB indicators (the gap between the cut-offs obtained in the two conditions ranges from 2.37 to 4.50 points). These supplementary results allow us to confirm the relevance of the cut-offs obtained including the six PB indicators reported above. Full results obtained by excluding the biological measure of stress from the PB indicators can be found in Supplemental Materials, Table S8.

Discussion

The current study aimed to determine clinical cut-offs for the two most widely used validated PB questionnaires: the PBI and the PBA. Based on a combination of data coming from self-reported PB measures, PB clinical judgments, and a biological measure of chronic stress, our analyses yielded the following clinical cut-offs: a score of 74.6 for the PBI and 86.3 for the PBA. Such values were the most conservative that emerged from the analysis strategy, in particular from the analysis of the PBI and PBA means of parents classified as cases, i.e., suffering from impairing PB levels. Less conservative cut-off values resulted from the ROC analyses: 53.7 for

the PBI and 52.7 for the PBA. As explained in the preregistration, we suggest using the most conservative values (i.e., 74.6 and 86.3 for the PBI and PBA, respectively) as clinical cut-offs that indicate the presence of the most severe levels of PB.

However, while most cut-off research leads to a single cut-off score distinguishing between patients with and without the condition under study, we would suggest employing the less conservative values (i.e., 53.7 and 52.7, respectively) as risk indicators of moderate PB severity. Such cut-offs should be used to signal parents for whom a more in-depth assessment is needed to ascertain if their suffering requires attention, support, or a specific intervention for PB even if they do not reach the highest threshold. Indeed, the less conservative cut-offs inform us that also parents with lower PBI and PBA scores may feel the need to seek treatment, perceive their distress as serious and impairing, and be deemed to be in great suffering by expert judges. We cannot exclude that low scores on questionnaires may be the effect of social desirability: shame and fear of judgment that often accompany the experience of burned-out parents (Hubert & Aujoulat, 2018) may lead them to underreport their symptoms (Roskam et al., 2017).

The idea to use the most conservative values as clinical cut-offs stems from our will to avoid overdetection of the PB condition (i.e., minimizing false positives). The suggestion to use the less conservative values as risk indicators of moderate PB severity aims to avoid suffering parents with lesser symptoms going unnoticed (i.e., minimizing false negatives). From this perspective, the current findings provide practitioners with two useful benchmarks to facilitate identifying not only parents suffering from the most severe levels of PB but also parents suffering from lesser levels who may need help. Classifying parents into one of these severity ranges based on their PBI and/or PBA scores may be highly informative as a starting point for assessment and treatment decisions.

Importantly, the severity ranges do not undermine the notion of PB as a continuum nor the use of continuous scores. It is preferable to treat PB as a continuum in most research situations. And even in clinical settings, the very score of the parent will always be more informative than a simple category. Indeed, a parent with a PBA score of 3 is clearly not comparable to a parent with a score of 51, even if they both fall into the "not in PB" category according to the above-mentioned cut-offs. And a parent with a score of 87 is not comparable to a parent with a score of 130, even if they both fall into the "suffering from the most severe PB levels" category. Furthermore, cut-off scores should not be used as the sole criterion while assessing potentially burned-out parents. The assessment and treatment planning should integrate results from multiple psychological tests and clinical interviews and consider differential diagnoses (e.g., job burnout and depression; Mikolajczak et al., 2020) to obtain a complete clinical picture of the parent and ascertain the sources of their suffering.

Without undermining the value of continuous scores and comprehensive assessment, these cut-offs nevertheless make an important contribution to the PB literature on account of their implications for identifying, treating, and preventing this condition. As regards implications for identification and treatment, the highest cut-offs provide practitioners and researchers with clear clinical benchmarks. Clinicians in search of cut-offs to validate their suspicion of PB can now rely on more stringent and reliable cut-off scores than the values used in previous studies, which were arbitrarily chosen and based solely on self-reported information. These arbitrary cut-offs can now be replaced by more rigorous values resulting from the implementation of a multi-informant and multi-method approach, whereby the limitations of one criterion are offset by the strengths of the others (Hajian-Tilaki, 2013).

Beyond their usefulness in clinical settings, cut-off scores were also awaited by PB researchers to move a step forward in the investigation of PB prevalence. Prevalence—the proportion of people affected by a particular condition in a given location at a particular time—is among the most fundamental measures in epidemiology. Prevalence estimates are indeed essential for three reasons (Ward, 2013). First, they are indispensable for public health policy decision-making: a high number of parents with burnout warrants preventive actions (whose success will be appraised by comparing prevalence estimates) and requires training and staffing more professionals than a low number. Second, they provide a useful context for clinical decision-making: if PB is more common than, say, myalgic encephalomyelitis, it would be useful information in evaluating a stay-at-home mother describing intense fatigue. Third, prevalence estimates are important to compare disease burden across locations or time periods. Without prevalence estimates, it is difficult to judge if PB is a growing, stable, or decreasing phenomenon. In order to progress on the issue of prevalence, clinical cut-offs were needed, and the current study constitutes an important step to this end.

The current study represents the first systematic attempt to derive reliable cut-off scores for PB. Despite its strengths, it is not without limitations. The preregistered methodology employed in this study has the strengths of having been conceived by a pool of 13 experts and based on a careful literature review on the cut-off determination. This methodology attempted to offer a rigorous procedure to deal with the lack of benchmarks in this domain. Future investigations should, however, verify the relevance of the cut-offs obtained in the current study in other samples and contexts and investigate their predictive properties in the short and long terms. This would be of particular relevance for the cut-offs that constitute risk indicators of moderate PB severity as it would allow testing of whether and how likely parents who suffer

from lesser PB levels tend to develop more severe PB levels, with critical clinical implications. Furthermore, in our analysis strategy, we anticipated the possibility of obtaining different cut-offs, which led us to set guidelines on the interpretation of more and less conservative cut-offs. However, this strategy would have been more robust if we had set an algorithm *a priori* with specific requirements in terms of sensitivity and specificity (for instance, select the cut-off values that produce the highest specificity, without decreasing sensitivity below .70 for any criterion). If we had the opportunity to redo the study from scratch, we would add this important step.

Second, results observed on the PB biological indicator suggest that HCC might not be able to discriminate between higher and lower levels of PB symptoms as assessed via the PBI and PBA. In the preregistration phase, we had sound reasons to believe in the usefulness of considering HCC among PB indicators (Brianda, Roskam, & Mikolajczak, 2020; Brianda, Roskam, Gross, et al., 2020). Yet, the sensitivity and specificity of the cut-off values obtained from the HCC indicator were very low, as was the PBI and PBA ability to discriminate between cases vs. controls according to this indicator. One possible explanation may lie in the fact that cortisol concentrations also - or mainly - depend on factors other than PB (e.g., other sources of stress or exhaustion). Future studies are needed to address the convergence between the physiological and psychological levels of PB (Semmer et al., 2003) and clarify the utility of considering HCC as a biological marker of PB.

Finally, a further limitation concerns the homogeneous nature of the sample and the presence of unequally distributed variables. Participants in this study were predominantly women with medium to high levels of education, partnered, and employed. The underrepresentation of fathers and the high percentage of parents with a university degree or higher has been observed in several studies in the PB field (see, e.g., Mikolajczak et al., 2019; Mikolajczak et al., 2018).

Moreover, the very small number of unemployed parents in the total sample (only 38 out of 192 participants) and the almost complete absence of unemployed parents in the subsample recruited as a control (only 2 parents) represent a major sampling limitation of the present study. The underrepresentation of unemployed participants has already been observed in the PB literature (see, e.g., Lindström et al., 2011; Roskam et al., 2021). As for the findings of the present study, it does not allow us to ascertain the relevance of the cut-offs for all parents regardless of their employment status. The results of the additional analyses performed on the homogeneous sample of employed parents would suggest that work status was not a confounding variable in the determination of the cut-offs in this study. However, to definitively rule out the hypothesis, the same analyses would have to be replicated on a homogeneous sample of unemployed parents (which was impossible with our data because of the aforementioned very small number of unemployed parents). Therefore, future studies should verify the relevance and generalizability of the cut-offs determined in the current study with more representative samples of case and control parents, and notably with a more representative number of unemployed parents. Such studies could, for instance, implement alternative methods to self-selection for the participants' recruitment [see, e.g., the Aunola and colleagues' (2020) recruitment strategy to increase fathers' participation].

Endnotes

[1]Several of these validation papers have already been published and others are currently in press in a special issue devoted to the measurement of parental burnout in New Directions in Child and Adolescent Development (https://onlinelibrary.wiley.com/doi/10.1002/cad.20286).

[2]The categories of "moderate" and "severe burnout" were chosen *a priori* and agreed upon by the team of 13 experts who participated in the design of the analysis procedure as both worthy of clinical attention.

[3]The threshold of seven for classifying PB levels was the result of an *a priori* choice agreed upon by the 13 experts. Scores equal to or greater than seven would allow for identifying parents who fall in the upper portion of the distribution for their PB level and thus most likely suffering from higher degrees of PB.

[4]This value was identified by Manenschijn et al. (2012) as the threshold that distinguishes people with a medical hypercortisolism condition from healthy controls. In their study, HCC was measured using the same analytical procedure as that used in the present study. Given the lack of validated normative references for HCC in burnout literature, we chose to rely on this value as the most relevant to identify clinically significant HCC.

[5]The preregistered procedure resulted from the joint work of 13 experts. First, the principal investigators of the study (first, second, and last author) conceived the procedure based on a review of the existing literature on cut-off determination in clinical psychology, as well as on their expertise in the field of PB. After the completion of the data collection, the procedure was sent to the ten co-authors (i.e., the pool of eight judges and two additional external researchers,

599 both experts in PB, who did not take part in the evaluations or other previous steps of the study)
600 for revision and approval.
601

References

Akobeng, A. K. (2007). Understanding diagnostic tests 3: Receiver operating characteristic curves. *Acta Paediatrica, International Journal of Paediatrics*, *96*(5), 644–647. https://doi.org/10.1111/j.1651-2227.2006.00178.x

Alexander, L. A., McKnight, P. E., Disabato, D. J., & Kashdan, T. B. (2017). When and How to use Multiple Informants to Improve Clinical Assessments. *Journal of Psychopathology and Behavioral Assessment*, *39*(4), 669–679. https://doi.org/10.1007/s10862-017-9607-9

Ambady, N. (2010). The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry*, *21*(4), 271–278. https://doi.org/10.1080/1047840X.2010.524882

Augenstein, T. M., Thomas, S. A., Ehrlich, K. B., Daruwala, S., Reyes, S. M., Chrabaszcz, J. S., & De Los Reyes, A. (2016). Comparing Multi-Informant Assessment Measures of Parental Monitoring and Their Links With Adolescent Delinquent Behavior. *Parenting*, *16*(3), 164–186. https://doi.org/10.1080/15295192.2016.1158600

Aunola, K., Sorkkila, M., & Tolvanen, A. (2020). Validity of the Finnish version of the Parental Burnout Assessment (PBA). *Scandinavian Journal of Psychology*. https://doi.org/10.1111/sjop.12654

Baldisserotto, M. L., Theme-Filha, M. M., Griep, R. H., Oates, J., Renó Junior, J., & Cavalsan, J. P. (2018). Transcultural adaptation to the Brazilian Portuguese of the Postpartum Bonding Questionnaire for assessing the postpartum bond between mother and baby. *Cadernos de Saude Publica*, *34*(7). https://doi.org/10.1590/0102-311X00170717

Brianda, M. E., Roskam, I., Gross, J. J., Franssen, A., Kapala, F., Gérard, F., & Mikolajczak, M. (2020). Treating Parental Burnout: Impact of Two Treatment Modalities on Burnout Symptoms, Emotions, Hair Cortisol, and Parental Neglect and Violence. *Psychotherapy and*

Psychosomatics, *89*(5), 330–332. https://doi.org/10.1159/000506354

Brianda, M. E., Roskam, I., & Mikolajczak, M. (2020). Hair cortisol concentration as a biomarker of parental burnout. *Psychoneuroendocrinology*, *117*. https://doi.org/10.1016/j.psyneuen.2020.104681

Caparros-Gonzalez, R. A., Romero-Gonzalez, B., Strivens-Vilchez, H., Gonzalez-Perez, R., Martinez-Augustin, O., & Peralta-Ramirez, M. I. (2017). Hair cortisol levels, psychological stress and psychopathological symptoms as predictors of postpartum depression. *PLoS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0182817

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*(1), 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617

Demetriou, C., Ozer, B. U., & Essau, C. A. (2015). Self-Report Questionnaires. In *The Encyclopedia of Clinical Psychology* (pp. 1–6). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118625392.wbecp507

Fluss, R., Faraggi, D., & Reiser, B. (2005). Estimation of the Youden Index and its associated cutoff point. *Biometrical Journal*, *47*(4), 458–472. https://doi.org/10.1002/bimj.200410135

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635. /pmc/articles/PMC3755824/?report=abstract

Hubert, S., & Aujoulat, I. (2018). Parental burnout: When exhausted mothers open up. *Frontiers in Psychology*, 9. https://doi.org/10.3389/fpsyg.2018.01021

Kawamoto, T., Furutani, K., & Alimardani, M. (2018). Preliminary validation of Japanese version of the parental burnout inventory and its relationship with perfectionism. *Frontiers*

in Psychology, *9*. https://doi.org/10.3389/fpsyg.2018.00970

Lindström, C., Åman, J., & Norberg, A. L. (2011). Parental burnout in relation to sociodemographic, psychosocial and personality factors as well as disease duration and glycaemic control in children with Type 1 diabetes mellitus. *Acta Paediatrica, International Journal of Paediatrics*, 100(7), 1011–1017. https://doi.org/10.1111/j.1651-2227.2011.02198.x

Magaña, A. B., Goldstein, M. J., Karno, M., Miklowitz, D. J., Jenkins, J., & Falloon, I. R. H. (1986). A brief method for assessing expressed emotion in relatives of psychiatric patients. *Psychiatry Research*, *17*(3), 203–212. https://doi.org/10.1016/0165-1781(86)90049-1

Manenschijn, L., Koper, J. W., van den Akker, E. L. T., de Heide, L. J. M., Geerdink, E. A. M., de Jong, F. H., Feelders, R. A., & van Rossum, E. F. C. (2012). A novel tool in the diagnosis and follow-up of (cyclic) Cushing's syndrome: measurement of long-term cortisol in scalp hair. *The Journal of Clinical Endocrinology and Metabolism*, *97*(10), E1836-43. https://doi.org/10.1210/jc.2012-1852

Maslach, C., Jackson, S. E., Leiter, M. P., Schaufeli, W. B., & Schwab, R. L. (1986). *Maslach burnout inventory*. Consulting psychologists press.

Mikolajczak, M., Brianda, M. E., Avalosse, H., & Roskam, I. (2018). Consequences of parental burnout: Its specific effect on child neglect and violence. *Child Abuse and Neglect*, *80*, 134–145. https://doi.org/10.1016/j.chiabu.2018.03.025

Mikolajczak, M., Gross, J. J., & Roskam, I. (2019). Parental Burnout: What Is It, and Why Does It Matter? *Clinical Psychological Science*, 216770261985843. https://doi.org/10.1177/2167702619858430

Mikolajczak, M., Gross, J. J., Stinglhamber, F., Lindahl Norberg, A., & Roskam, I. (2020). Is

671 Parental Burnout Distinct From Job Burnout and Depressive Symptoms? *Clinical*
672 *Psychological Science*, 216770262091744. https://doi.org/10.1177/2167702620917447
673 Mikolajczak, M., & Roskam, I. (2018). A theoretical and clinical framework for parental
674 burnout: The balance between risks and resources (BR2). *Frontiers in Psychology*, *9*.
675 https://doi.org/10.3389/fpsyg.2018.00886
676 Park, S. H., Goo, J. M., & Jo, C. H. (2004). Receiver operating characteristic (ROC) curve:
677 Practical review for radiologists. *Korean Journal of Radiology*, *5*(1), 11–18.
678 https://doi.org/10.3348/kjr.2004.5.1.11
679 Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained
680 using two criteria based on the receiver operating characteristic curve. *American Journal of*
681 *Epidemiology*, *163*(7), 670–675. https://doi.org/10.1093/aje/kwj063
682 Pintea, S., & Moldovan, R. (2009). The Receiver-Operating Characteristic (ROC) analysis:
683 Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral*
684 *Psychotherapies*, *9*(1), 49–66.
685 Roskam, I., Aguiar, J., Akgun, E., Arikan, G., Artavia, M., Avalosse, H., Aunola, K., Bader, M.,
686 Bahati, C., Barham, E. J., Besson, E., Beyers, W., Boujut, E., Brianda, M. E., Brytek-
687 Matera, A., Carbonneau, N., César, F., Chen, B.-B., Dorard, G., … Mikolajczak, M. (2021).
688 Parental Burnout Around the Globe: a 42-Country Study. *Affective Science*, 1–22.
689 https://doi.org/10.1007/s42761-020-00028-4
690 Roskam, I., Brianda, M. E., & Mikolajczak, M. (2018). A step forward in the conceptualization
691 and measurement of parental burnout: The Parental Burnout Assessment (PBA). *Frontiers*
692 *in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00758
693 Roskam, I., & Mikolajczak, M. (2020). Gender Differences in the Nature, Antecedents and

694 Consequences of Parental Burnout. *Sex Roles*, 1–14. https://doi.org/10.1007/s11199-020-
695 01121-5
696 Roskam, I., Raes, M. E., & Mikolajczak, M. (2017). Exhausted parents: Development and
697 preliminary validation of the parental burnout inventory. *Frontiers in Psychology*, *8*.
698 https://doi.org/10.3389/fpsyg.2017.00163
699 Russell, E., Kirschbaum, C., Laudenslager, M. L., Stalder, T., De Rijke, Y., Van Rossum, E. F.
700 C., Van Uum, S., & Koren, G. (2015). Toward standardization of hair cortisol
701 measurement: Results of the first international interlaboratory round robin. *Therapeutic*
702 *Drug Monitoring*, *37*(1), 71–75. https://doi.org/10.1097/FTD.0000000000000148
703 Semmer, N. K., Grebner, S., & Elfering, A. (2003). Beyond self-report: Using observational,
704 physiological, and situation-based measures in research on occupational stress. In P. . L.
705 Perrewe & D. C. Ganster (Eds.), *Research in Occupational Stress and Well Being* (Vol. 3,
706 pp. 205–263). Emerald Group Publishing Limited. https://doi.org/10.1016/S1479-
707 3555(03)03006-3
708 Stalder, T., & Kirschbaum, C. (2012). Analysis of cortisol in hair--state of the art and future
709 directions. *Brain, Behavior, and Immunity*, *26*(7), 1019–1029.
710 https://doi.org/10.1016/j.bbi.2012.02.002
711 Staufenbiel, S. M., Penninx, B. W. J. H., Spijker, A. T., Elzinga, B. M., & van Rossum, E. F. C.
712 (2013). Hair cortisol, stress exposure, and mental health in humans: A systematic review.
713 *Psychoneuroendocrinology*, *38*(8), 1220–1235.
714 https://doi.org/10.1016/j.psyneuen.2012.11.015
715 Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver
716 operating characteristics curves. *Canadian Journal of Psychiatry*, *52*(2), 121–128.

https://doi.org/10.1177/070674370705200210

Van Bakel, H. J. A., Van Engen, M. L., & Peters, P. (2018). Validity of the Parental Burnout Inventory among dutch employees. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00697

Ward, M. M. (2013). Estimating disease prevalence and incidence using administrative data: Some assembly required. *Journal of Rheumatology*, *40*(8), 1241–1243. https://doi.org/10.3899/jrheum.130675

Table 1

Total and Percentage of Parents Classified as Cases (i.e., Most Likely Suffering From Impairing PB Levels) According to the six Indicators

Indicator	Total	Parents classified as cases	
		n	*%*
a (seeking PB treatment)	192	130	67.7%
b (self-reported PB category)	191	93	48.7%
c (self-reported PB levels)	184	87	47.3%
d (judges-reported PB category)	107	51	47.7%
e (judges-reported PB levels)	107	36	33.6%
f (hair cortisol concentration)	183	56	30.6%

Notes. *a* = parents voluntarily enrolled in PB treatment; *b* = parents self-reporting a PB category at least moderate; *c* = parents self-reporting PB levels of at least seven on a scale from zero to 10; *d* = parents to whom at least one judge attributed a PB category at least moderate; *e* = parents to whom clinical judgments attributed a PB level of at least seven on a scale from zero to 10; *f* = parents with hair cortisol concentration greater than or equal to 75.9 pg/mg hair (see the Statistical analyses section).

Table 2

Area Under the Curve (AUC), 95% Confidence Intervals (CI), and p-values of PBI and PBA Ability to Discriminate Between Cases and Controls According to the Classifications of the Six PB Indicators

Test	Indicator	*N*		AUC	95% CI	*p*
		Cases	Controls			
PBI	*a* (seeking PB treatment)	124	62	0.97	0.94-0.99	<0.001
	b (self-reported PB category)	89	96	0.88	0.84-0.93	<0.001
	c (self-reported PB levels)	83	95	0.90	0.85-0.94	<0.001
	d (judges-reported PB category)	47	55	0.96	0.93-0.99	<0.001
	e (judges-reported PB levels)	33	69	0.91	0.85-0.97	<0.001
	f (hair cortisol concentration)	54	125	0.58	0.50-0.67	0.09
PBA	*a* (seeking PB treatment)	130	62	0.98	0.96-1.00	<0.001
	b (self-reported PB category)	93	98	0.90	0.85-0.94	<0.001
	c (self-reported PB levels)	87	97	0.91	0.87-0.95	<0.001
	d (judges-reported PB category)	51	56	0.97	0.95-1.00	<0.001
	e (judges-reported PB levels)	36	71	0.94	0.89-0.98	<0.001
	f (hair cortisol concentration)	56	127	0.59	0.50-0.67	0.06

Notes. PBI = Parental Burnout Inventory. PBA = Parental Burnout Assessment

Table 3

Cut-Off Values for Each Indicator, Sensitivity, Specificity, Average Cut-Offs, and CIs Resulted From the ROC Analysis and the Analysis of Means

Indicator	PBI							PBA						
	ROC analysis			Analysis of means				ROC analysis			Analysis of means			
	Cut-off value [a]	SE (%)	SP (%)	*M*	*SD*	SE (%)	SP (%)	Cut-off value [a]	SE (%)	SP (%)	*M*	*SD*	SE (%)	SP (%)
a (seeking PB treatment)	41.50	88.7	95.2	73.98	24.57	56.7	98.4	34.00	92.3	93.5	84.38	30.46	54.6	100
b (self-reported PB category)	64.50[b]	80.9	82.3	79.00	23.72	51.7	90.6	61.00	87.1	80.6	91.54	28.69	55.9	91.8
c (self-reported PB levels)	65.50	81.9	85.3	79.83	25.40	54.2	94.7	63.50[b]	87.4	82.5	93.22	29.15	58.6	94.8
d (judges-reported PB category)	41.50	83.0	94.5	73.85	29.17	59.6	100	30.00	94.1	92.9	85.20	32.80	52.9	100
e (judges-reported PB levels)	73.50[c]	72.7	94.2	79.39	29.64	60.6	94.2	63.00	88.9	88.7	93.97	30.62	58.3	97.2
f (hair cortisol concentration)	35.50[b]	75.9	41.6	61.44	30.01	55.6	53.6	64.50[b]	60.7	55.9	69.25	39.37	57.1	56.7

Average cut-off	53.67	74.58	52.67	86.26
95% CI	41.99 - 65.34	69.48 – 79.68	40.91 – 64.43	79.49 – 93.03

Notes. PBI = Parental Burnout Inventory. PBA = Parental Burnout Assessment. SE = sensitivity. SP = specificity.
[a]PBI and PBA scores associated with the combination of the highest sensitivity and highest specificity [when the values reported in this column are not superscripted, it means that Youden's index and the Closest-to-(0-1) criterion yielded the same score]
[b]Determined using the Closest-to-(0,1) criterion because it led to the most conservative values
[c]Determined using the Youden index because it led to the most conservative values