How to make PhD students graduate on time

Dr. Joni Salminen, University of Vaasa
Email: jonisalm@uwasa.fi
June 14, 2024 (V3[1])

Foreword: Written to educators. Contains my opinions: use them if you want, don't use them if you don't want/like/agree or whatever. My intention is good: writing this to help. I believe there are serious flaws in how PhDs are supervised in Finland (and possibly elsewhere), and I'm outlining here some principles I believe would help fix the problems.

Acknowledgment: Most if not all of this I've learned under the supervision of Dr. Bernard J. Jansen while working at Qatar Computing Research Institute (QCRI). A huge debt of gratitude to him, as he basically taught me everything I know about making research happen.

…so, four steps to making this work. You must remember PhD students come in as "blank pages"; they have very little experience. So, you need to set the right expectations and hold their hand, if you want them to graduate quickly.

Or, you can let them figure out the things by themselves, impose no rules, and not ask them to report back to you all that time. But, that will get you SLOW graduation, which is probably what you want to avoid since you're reading this.

Here are the steps:

Step 0: Make sure YOUR research skills are updated.

This is zero, before anything starts. Before anything starts, you need to ensure you have the skills yourself, to accommodate a wide range of methods/approaches. Otherwise, it's the "blind leading the blind". In my opinion, at least the following should be known by anyone who has a PhD and is supervising PhD candidates:

- Experiment design (basics of science)
- How to argue for a research gap
- How to create good research questions
- How to structure research papers (+ an idea of how long each section should approximately be, though this varies depending on the venue. But an overall good length for a paper in business / HCI would be ~8000 words.)
- Basic statistics:
    - **Descriptive stats or EDA (exploratory data analysis)** --- means, standard deviations (to get an idea of variable tendencies and dispersion), histograms (to get an idea of variable distributions).
    - Understanding how **data variability** (=cell size for categorical data and dispersion for numerical data) affects test selection; understanding how **variable types** (categorical, numerical, ordinal, binary/multiclass) affect test selection.
    - **APA style reporting of statistics** (easy to google every time! But need to know that there is this convention).
    - **Staying organized** (declarative naming of EVERYTHING: datasets, variables, working files; "one sheet to rule them all" mentality to avoid creating new files all the

---

time; keeping all files of a given paper in the same local or cloud folder; backing up everything all the time).

- o **Having some grasp of rules of thumb** (this is a bit tricky and statisticians disagree on these, too, but things like "minimum cell size per categorical variable level = 5", "minimum sample size = 30", "max. levels to analyze for categorical variable = 5"; again, these are my rules of thumb and they are not universally applicable. But a good supervisor has some rules of thumb to offer in order to alleviate the complexity of data analysis).
- o **Understanding of data dimensionality** (that sometimes data can be too sparse to get a signal and that combinations of variable levels generate more parameters to estimate which requires more observations, i.e., higher sample sizes)
- o **Tests that get you a long way:**
  - **T-test** (comparing means between two groups --- knowing this is good because it'll instill in you an important principle for research: comparing groups. So, you can come up with different research designs more easily)
  - **Mann-Whitney** (like previous but with non-normal data)
  - **ANOVA** (comparing means between more than two groups)
  - **Kruskal-Wallis** (like previous but with non-normal data)
  - **Chi-squared test of independence** (comparing counts between groups under different conditions; most typical is 2x2 arrangement; e.g., male/female; success/failure and the cells show frequency → there are easy-to-use online tools to run the test)
  - **Correlation/regression** (examining how multiple numerical variables are associated with one another; can get a p-value out of correlation, too)
- o **Difference between omnibus and post-hoc tests** (ANOVA/Kruskal-Wallis are omnibus; if they show significant differences, post-hoc tests follow)
- Systematic literature reviews (everyone in research needs to know this)
- Qualitative coding of data (everyone in research needs to know this – also, you *can* (and should) analyze qualitatively coded data in a quantitative way ---- now, you don't need to use a specialized tool like NVivo or Atlas.ti. I used NVivo in my dissertation but nowadays I only use Excel for coding! It's enough – I typically add new columns when hierarchical codes are needed. I typically use qualitative analysis to support and contextualize quantitative findings, which I also teach my students – this corresponds to typical way of doing mixed-methods research, at least in fields like human-computer interaction)
- (Applied ML) (in parentheses because may not be a requirement – but useful to know, especially for research in computational social sciences; this involves things like supervised/unsupervised ML, train-test paradigm, ML analysis steps, with associated best practices like replication/model saving).

(When it comes to statistics and ML, things are surprisingly simple if you keep some "best practices" and guiding principles in mind. You don't need to be an expert statistician to teach these. Of course, you should learn from expert statisticians and e.g. follow statistics on LinkedIn – there's a lot of depth to this field, but using it doesn't require complete understanding.)

Now, you could teach your PhD candidates only one type of method or approach, and force your student to use it, too (as many do). But PhD completion is not the only goal: the true and genuine goal is to train professional researchers (not one-trick ponies!). For this, multiple skills are needed – a person completing their PhD should know how to do a quantitative study, a qualitative study, a systematic review; they should learn how to design an experiment and how to analyze data

statistically and qualitatively. Also, they should have other skills like some teaching and supervision of others, but first and foremost they must learn the trade of research.

So, as a supervisor, to be able to teach your candidates hands-on, you need to have first-hand experience with this broad range of skills yourself. However, surprisingly many PhDs nowadays lack basic research skills, at least in Finland (and this forms a vicious cycle: bad method knowledge creates more PhDs with bad method knowledge who then teach more PhDs with bad method knowledge and suddenly it becomes accepted that you can have a PhD and don't even know how to design an experiment).

So, this is one reason why PhD completion is slow: your supervisor cannot teach you things, so you're lost at sea figuring out everything by yourself. This is slow, painful, and leads to a loss of a lot of potential. Supervisors need to first get their method skills in order. Again, crucial step because without this, none of the following steps will work.

Step 1: Selection

Set standards. I don't accept students who meet the following conditions:
- Unwillingness to dedicate at least one day per week to fully working on their research
- Unwillingness to give weekly reports
- Doesn't pass a screening test (more on this below)
- Works on a different field or topic than you – why would you take them? There's little win-win and you can't help them as much as you can with students working on your field and topic.

Notice that I didn't say "Don't take part-time PhD students". According to my experience, "part-time" is not a condition for failure. It all comes down to the student's maturity and ability to manage their time. A lazy full-time student can be much slower than an active part-time student, by A LOT. Especially people who have an entrepreneurship background can handle high pressure, steady progress on a part-time basis. They know how to manage their time and they are good at taking feedback (well, many of them at least). Entrepreneurs and former athletes make great PhD candidates because they are disciplined and can generally follow instructions. Doing a PhD is less about being smart and more about willingness to take feedback and learn from it. Also, not all PhD candidates need to become *top* researchers – that's completely fine. There is variation in how quickly students learn. Everyone can still get the PhD done if they keep trying, taking the feedback, improving, and – yeah, just keeping at it.

Apply a screening task. Why?: everyone sounds good on the phone. Most people can write sweet things in their emails. Students promise you the moon, and their CVs might look okay. But do they know anything? Are they actually willing to work? Finding out is easy. Give them a screening task. Appendix 1 has an example of a screening task I use. It tests two things: (1) is the person able to follow instructions, (2) does the person show creative abilities, and (3) is the person motivated enough to get the job done.

Interestingly enough, most candidates never return the task! So, you dodged a bullet there. Those that do, assess if their output is good enough. If not, politely refuse them. "Good enough" doesn't mean it's ready – it means it can be developed into usable. This is a judgment call.

…for some reason – and I really can't understand why – doctors in Finnish universities tend to take into their supervision students from a totally random topic not related to your own field or topic.

Cannot understand why. As a supervisor, you're spending time on someone who is not part of contributing to your own research agenda. You should not spend mental energy figuring out new topics in depth. I'd say: don't do it. Say no.

Maybe the reason is that they get fairly few good candidates so the ones that are promising they take, regardless of topic match – but, this just perpetuates the situation:
take poorly matching candidates → *get poor results* → *get more poorly matching candidates*
versus the better idea:
take well matching candidates → *make them work under you in topics where you want to be the best* → *become more known in those topics* → *get more matching candidates* → *etc.*

One more thing: don't take too many PhD students. Here's another mistake I see: a person has under their supervision multiple students (say, 12). But in name only. The supervisor doesn't even know the students' current status! It's like having a lot of visitors on an e-commerce website of whom nobody converts: useless traffic. In my approach, I'm giving the students a lot of my time and cognitive energy, so I need to restrict their numbers. The idea is to have fewer students at once but to get those that I have done faster. This is better than having 12 students with whom you're making no real progress. I would say 3 students is already at the limit of one supervisor's capabilities, considering that we have other duties as well. But it is very simple: quality supervision takes time. Quality is more important than quantity. So, take fewer students and give those you take more time.

Step 2: "Blood, sweat, tears"

I use Churchill's phrasing to set expectations to my prospective candidates: *Blood, sweat, tears*. I tell prospective students that working with me is difficult. It's more demanding than working with an average professor. I give direct feedback that might hurt their feelings. I require dedication, loyalty, and commitment. I don't promise them a lot of nice things, like freedom and time. In fact, I promise them the opposite: control and hurry.

…for some reason, few candidates have ever rejected me as a supervisor because of this. On one hand, this can be because screening works, and the candidates I talk to end up being serious. On the other hand, it might be that students actually like rigor. They want you to be strict, so that they learn more effectively. They also want to graduate soon and get on with their lives – they just don't know how. So, they tend to accept tighter control, because it benefits them.

Of course, in exchange for their dedication, I promise them high-quality support (that's about the only thing I promise – "avoid making any promises" is a good mantra in professional life). This one promise is a must. I also tell that they're not only getting me, but they become part of a research team. This is crucial. You, as a supervisor, should have a research team. If not, you should build one. "One-person research" is doomed to remain a struggle – you need multiple talented people. A proper team can support the PhD student more than you could alone.

So, anyway, the expectation is: with me, it's all about hard work. But you get a chance of getting good results out of it.

Step 3: Support, support, support

This is perhaps the part Finnish universities most struggle with. The typical way is this: "Go out and do something. Come back when you have something." As a student, you're thinking "…okay." There are a million thoughts in your head, ranging from topics to methods to writing styles and theories. It's

way too messy, way too complex (it's Alice in Wonderland, jumping to the rabbit hole and getting into things she shouldn't get into). And to make it even messier and more complex, the supervisor says, "start with doing a literature review". So, now the student spends the first six months reading on material that does *not* make things simpler, but it adds to the complexity: more ideas, more approaches, more alternatives. So, as a supervisor, I want to do two things: (1) take away options (=complexity) and (2) get the student quickly past the cold start problem of having no data to analyze. Without data, there is no research. As soon as we have that sweet data, we have now a paper in our hands; we just need to make it happen. But if we read literature, then we read literature. We are not doing research.

So, I tell exactly the opposite: "we will plan the first study together. Don't read ANYTHING except the papers I send to you." If I do send them papers, it's a few papers, max. It's a huge, huge risk to send a PhD student into the rabbit hole of literature. It's especially dangerous in the beginning: it'll get them confused, 100%. So, start very small and start by explaining to them the research gaps. You should know, you are the supervisor, and you're a professional researcher – remember, you chose a student who is working in the same field as you. So, you can help them out readily. Another thing here is that almost any study can be rationalized ex post; a gap is easy to justify and explain, once you know how to do it. And anyway, as I said, if you've followed this guide, the student is working on your topical area and you already know the gaps, so you won't risk misleading them. They will learn the gap identification throughout their PhD journey, as they start taking on more responsibilities over the course of subsequent papers.

But in the beginning, focus on action is needed. Believe me, candidates will find the complexity from the literature at any time. That lesson isn't hard to learn. To get them going, get them as soon as possible to a stage where they can collect the data. You need to help them a lot with RQs, hypotheses (if you have any; I won't necessarily even recommend having hypotheses but just RQs) and basically everything. So, give detailed instructions, do the heavy lifting together with them, so they see how you do it and can start mimicking. The good old "master and disciple" approach!

I've seen this work and found some theoretical basis for it as well[2]. You do it a couple of times and within a year, the student develops enough skills to take on more and more areas; eventually, they'll be able to do the whole process by themselves. But it's much faster to show them than making them read it from some books. "Learning by doing", it's no more special than that. In fact, it's so simple that there's no excuse not to do it like that, in my books. If you have selected the right candidates, you don't have too many students, and you have the skills, why not show the student how to do it?

Here are some parting principles:

- **Teach the student hands-on skills:** On my first day at QCRI, Jim sat down with me and showed how to preprocess data in Excel. Forever blew my mind. A professor can actually show you things, not just say, "Figure it out". Ever since then, I've been passing on this lesson.

---

[2] Direct quote (**bolding** by me): "Wood et al.'s [73] theory of **scaffolding** emerged around 1976 as a part of social constructivist theory and was particularly influenced by the work of the Russian psychologist Lev Vygotsky [69]. Vygotsky [69] argued that we learn best in a social environment, where we construct meaning through interaction with others. His **Zone of Proximal Development (ZPD)** Theory, which states that we can learn more in the presence of a knowledgeable other person, became the basis for the Theory of Scaffolding [22]. Among other things, this theory states that when learners strive to acquire new knowledge, they need individualized support falling within their individual ZPD. **ZPD represents the potential distance the learner could reach with the help of a more knowledgeable other** [69]. **As they advance and become more independent in their thinking, this support can gradually fade away.**" Source: Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3313831.3376781

They need to write RQs? Show them how to do it. They need to collect data? Show them how to do it. They need to analyze it qualitatively? Show them how to do it. Quantitatively? Show them. They need to write a paper outline? Show them how to do it. And so on, until every little thing is understood by them. Takes time? Yes. That's why, in Step 1, we don't take many students. Requires you to know your stuff? Yes. That's why in Step 0, we cannot be good supervisors if we don't know how to do things.

- **Monitor.** For example, require weekly meetings or email reports. In the beginning, higher meeting frequency is required, because the student is basically starting from scratch in terms of skill and substance. They will need a lot of handholding, and you're the person for the job.
- **Stay organized.** For example, set up a Trello board with all task items (see Appendix 2 for an example). This is useful because there are many papers to work on and many tasks within a paper. So, staying organized is crucial.
- **Organize funding.** You as the supervisor should bear the responsibility of funding your PhD students. You don't want them to feel uncertainty, or waste time on writing funding applications (that'll come later in their careers). You want to create a safe environment for them to focus on learning how to do research. So, using whatever means you have, get that funding.
- **Create a culture of responsiveness.** Reply within minutes, not within days. Prioritize the emails coming from your PhD students and make sure you remove their roadblocks. Be very brief and unambiguous in your answers; write clear instructions. Vice versa, require they respond quickly back to you, and if they don't, send them reminders in a strict, short way to get them moving.
- **Create a culture of haste.** My philosophy to research is this: we don't have much time. In fact, we always have too little time. Our lives and careers are over before we can make a dent in science. So, every minute literally counts. I want to instill this way of thinking in my students. I don't want them to think of academia as a slow-paced, relaxing environment for "interesting" and "insightful" thinking. Because it's not. It's an environment for execution and implementation. It's an environment of extreme competition – it's a tough and ruthless game! …of course, maybe later in their careers, once they've gotten tenure, they could relax and engage in some of that insightful thinking. But to get to that level, one needs to act, a lot. So, I want them to appreciate my time but first and foremost their own time. Every unnecessary task, like course completion to get the credits, admin things, presenting in research seminars, etc., I tell them to either avoid or do with minimal effort. What they need to focus on is their research. Everything else will follow.
- **Communication.** "Firm. Fair. Friendly. (In that order.)" This is one of the many great quotes from Dr. Jim. Be firm with your students. Make them accountable. If they are falling behind, set them straight. Sometimes they need to feel uncomfortable. Growth doesn't come without pain. So, "support" doesn't mean just being nice – it means pushing the student when it's needed. But in a fair way. Don't expect them to be superhumans. Be reasonable. You should have a conception of how long a task takes and apply that target when imposing deadlines. If unexpected things happen, be flexible. Ultimately, all this is a game. It's a serious game, but still a game. So, show some mercy as well. Be conscious of your students' well-being (see Appendix 3 for my message to students relating to "the place called extreme"). Give them a lot of praise when they do well: encourage them to make things happen.

Final words: Demand a lot from your student, and yourself. Focus on getting the job done. Keep moving forward.

APPENDIX 1: Screening task example

[Name], here is a more detailed to-do for you to prepare. It's a general set of instructions I prepared for all students that want to collaborate with us. You can also follow this process.

=======================
Hi!

Here are instructions to prepare for our meeting.

TASK 1
Review the persona research papers indicated in this blog post: https://jonisalminen.com/the-definitive-list-of-persona-research-read-these-to-get-a-foundational-understanding-of-persona-research/

TASK 2
Pick one of the following topical areas AND create three research questions out of it. Don't know how to write research questions? Read this: https://www.monash.edu/library/help/assignments-research/developing-research-questions

Personas and Social Impact (Social Good) -- how can personas be of help in societal problems, including socio-technical dilemmas and wicked problems?
Explainable Data-driven Personas -- how can algorithmically generated personas be explained to persona users in an effective way?
Novel Customer Segmentation Algorithms and Other Methods -- how customer and user segmentation algorithms be "renewed" and improved?
Immersive Deepfake and Generative AI Personas -- what does AI technology, including Generative AI and deepfake technology, mean for persona creation and use?
Integrating Personas into Automated Marketing Systems -- how can algorithmically generated personas be integrated into marketing systems such as online advertising platforms to support marketers' decision making and automatic marketing optimization?
Metaverse Personas -- what implications does the Metaverse (VR, AR) technology have for user personas?
Interactive Persona Systems -- how can various stakeholders interact with user and customer personas in realtime to accomplish their professional goals?

Write the research questions below.

RQ1: ???
RQ2: ???
RQ3: ???

An example:
RQ1: Does a persona user's emotional intelligence have an effect on how the user perceives personas?

TASK 3

For each research question, write in 1-3 sentences how you're going to address it.

*To address RQ1, we...*
To address RQ2, we...
To address RQ3, we...

An example:
To address RQ1, we conduct an experiment in which designers use personas in a realistic work task scenario. We measure the designers' emotional intelligence using a validated psychometric scale and we measure the designers' perceptions of personas using a validated Persona Perception Scale. We then measure the effect of the designer's emotional intelligence on the different persona perceptions.

NOTE: Make sure to complete the three tasks before our meeting, so we can immediately focus on getting the research started.

APPENDIX 2: Example of a Trello set-up
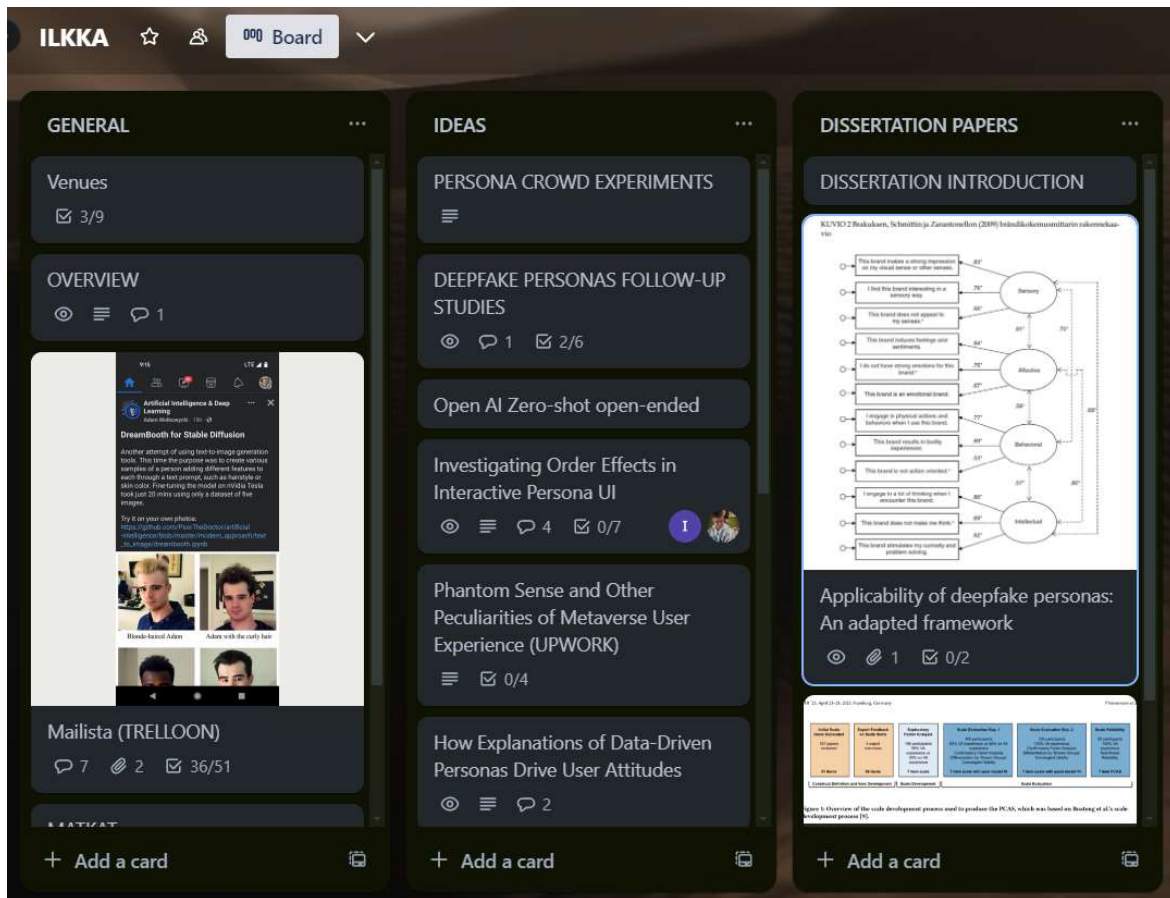
APPENDIX 3: On the place called extreme

Dear students,

you've all seen this place.
You know you're there when weird things start to happen: the research comes into your sleep.
You can't focus on anything else.

You feel stressed and on the edge.
You might have trouble eating, your stomach might have pain.
You might get a nosebleed.
(All of these have happened -- to me and to you, based on what you've told me.)

You're in the place called extreme: the in-between-area of creativity and obsession.

This place is sometimes necessary. Great ideas and outcomes can come from the visit. Like they say, diamonds are born under pressure. I want to highlight that this is different from burnout - in this place, you are very focused and very efficient. You might not notice your biological needs -- you're driven, you're focused, and you can make big leaps of progress with your research.

But the place is toxic: it's hazardous to your health.

You should visit it seldom and only for a short time. Only for days at a time (or *a* day), not weeks.

And the more mature you get, the less frequent should those visits become.

I'm writing this because you can't avoid this place completely.

Because you're PhD candidates or young post-docs, you have to push hard to overcome the many obstacles.

That has some price, but the price can't be too much.

Your health must come first.

Nobody else except you knows when you are the limit. Your supervisors, including me, can throw a lot of things at you. Unfortunately, we cannot see where your breaking point is. We of course think about it, and we don't want to overburden you. But we do want you to be *close* to the limit at this stage of your careers -- "close to limit" is the place where you grow.

"Close to limit" is okay. "The place called extreme" is not - it's a danger zone you need to just briefly visit when you have no other choice, and then quickly escape.

Some tactics that help here:
-learn to approach deadlines differently: instead of thinking, "Deadline is Friday, so I'll complete it on Friday", think "Deadline is Friday, so I'll complete it on *Monday*" or "Deadline is September, I'll complete it on May" (<= real approach in our team that has elevated our productivity while decreasing our stress)
-learn to say no to things --- it's tempting to say yes to things, especially at early stage of your career when you're looking for opportunities. Every opportunity, even if it doesn't really advance your career, seems like worth taking. You're eager to work. But the graveyard is full of careers of people who didn't learn how to say no! Unfortunately, there's no fixed rule on when you should say "no" and when "yes" -- sometimes, you make mistakes with both categories!
-keep moving forward. Got a rejection? Keep moving forward! Got another rejection? Keep moving forward! Didn't get the job? Keep moving forward. You get the picture. The people that endure are those that don't give up. But, there's an important nuance here - you *can* (and should) give up on things that are doomed. Projects and people that won't elevate you - get rid of those. What "not giving up" means is not giving up on the big picture, your "ultimate goal". Again, it's really hard to

distinguish between when to let something go and when to keep enduring. But with *some* things, you must endure, whatever they are.

-enjoy the journey: after you've reached your goal posts, be it PhD or Professorship, you'll notice yourself thinking, "Is this it? Is this all there is?". And you might take on some new, imaginary goal post to chase. We can never be really satisfied when we chase the goal posts because there's always another one in the horizon. So, it's more important to work with people who you enjoy working with --- it's much more fun and exciting. Try to find (or create) that team of people that'll stand by you through the good and the bad! The worse situation in academia is doing research alone. The best situation is doing research in a good team. Like I said to Waleed, "you'll KNOW when you get there."

Anyway, the main point here was about the place called extreme. Listen to your body and don't overdo it. Communicate clearly with your supervisor if you're overworked - I guarantee you that we'll understand. You don't need to suffer alone.

Best,
Joni