Subject: RE: multilevel modeling seems to be having a little resurgence these days
Date: Wednesday, August 24, 2022 at 9:59:12 AM Eastern Daylight Saving Time
From: Jay Kaufman, Dr.
To: Patricia O'Campo, Basile Chaix
CC: Sam Harper, Seungmi Yang

That is super interesting. This would be a great project in intellectual history, to trace these networks and connections and influences. There was clearly a split across the Atlantic, but I find it very interesting that on the US side, the center of gravity for application shifted from Michigan to Hopkins seemingly out of nowhere. Pat shows that the intellectual NEED for these methods was brewing in the early 1990s, and not just at Hopkins. But somehow Baltimore is where the theorists met the statistics and the software and it all came together. One of those magical moments. And Carles may be the mysterious patient 0. - JK

From: Patricia O'Campo <Patricia.OCampo@unityhealth.to>
Sent: Wednesday, August 24, 2022 1:32 PM
To: Basile Chaix <basile.chaix@iplesp.upmc.fr>; Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Sam Harper <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: RE: multilevel modeling seems to be having a little resurgence these days

Thanks Jay, appreciate tracing the origins in other disciplines. Before I attempted the stats, we (Carles and I) published a social epidemiological conceptual/theoretical paper calling for going beyond the individual (e.g., individualistic fallacy, see page 6 of the attached) and also moving beyond the proximal. I had wanted to model these relations but as you point out did not have statisticians who worked on this issue so I asked a friend (Wang) who lent me her Phd student (Xue) and we worked on both AJPH papers (1995, 1997) as I still didn't have access to HLM and wasn't aware of MLWin. I do think I might have read Entwistle's demography papers but I didn't dig deeper to see how she approached the stats. I'm certainly aware of her work. I took Raudenbush's course after my first paper and was only able to use a beta version of HLM but after I submitted my second paper. I did speak to Ana about MLM and Carles worked with her on that paper. So at Hopkins PH, Carles and I 'seeded' these ideas.

-Pat

From: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Sent: Wednesday, August 24, 2022 5:58 AM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Sam Harper, Dr. <sam.harper@mcgill.ca>; Patricia O'Campo <Patricia.OCampo@unityhealth.to>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

Thank you Jay for this impressive analysis conducted with the accuracy of a criminal investigation :)

Things are not always as new as we think. Ideas are floating in the airs and the most visionary (here Pat and Ana) can catch them.

I just remembered of the excellent review of G Kaplan on SES and cardiovascular diseases.
It was published in 1993, and I just double checked, Georges Kaplan had a specific section on area factors.
Here is the comment of Kaplan on one study: "In another, residence in a poverty area was

associated with increased risk of death from all causes, even when individual SES was taken into account".18
Although this study probably did not use multilevel analysis, most of the concept was already there.

Haan M, Kaplan GA, Camacho T.Poverty and health: prospective evidence from the Alameda County Study. Am J Epidemiol 1987;125:989-998.

The innovation around 1995-1997 was to incorporate multilevel models to answer types of questions that had already been raised in some studies.

Best regards,
Basile

De: "Basile Chaix" <basile.chaix@iplesp.upmc.fr>
À: "Basile Chaix" <basile.chaix@iplesp.upmc.fr>
Envoyé: Mercredi 24 Août 2022 12:46:15
Objet: TR: multilevel modeling seems to be having a little resurgence these days

---

De : Jay Kaufman, Dr. [mailto:jay.kaufman@mcgill.ca]
Envoyé : mercredi 24 août 2022 11:24
À : Sam Harper; Basile Chaix
Cc : Patricia O'Campo; Seungmi Yang
Objet : RE: multilevel modeling seems to be having a little resurgence these days

Makes sense to me that the first applications would be in sociology and criminology, since the first paper in the
citation chain is Mason and Wong 1985 (https://www.jstor.org/stable/2288464). Bill Mason was a prof of sociology at University of Michigan. I think that he was the PhD supervisor of Barbara Entwistle, as she went on to publish a whole bunch of sociology and demography papers using this methodology in the late 1980s, like:

Entwisle B, Hermalin AI, Kamnuansilpa P, Chamratrithirong A. A multilevel model of family planning availability and contraceptive use in rural Thailand. Demography. 1984 Nov;21(4):559-74.
Entwisle B, Mason WM, Hermalin AI. The multilevel dependence of contraceptive use on socioeconomic development and family planning program strength. Demography. 1986 May;23(2):199-216.
Entwisle B, Casterline JB, Sayed HA. Villages as contexts for contraceptive behavior in rural Egypt. American Sociological Review. 1989 Dec 1:1019-34.

As you know, Entwistle continued on to become chair of sociology at UNC and head of its pop center, and

married to Ken Bollen of SEM fame.

So I would say that sociology and demography (and thus criminology) were 10 years ahead of public health, and that education also got in early in this, since this was the interest of Goldstein in UK and also of Bryk and Raudenbush in US.

What I am curious about was the stimulus at Johns Hopkins for the contemporaneous innovations of Diez-Roux and O'Campo. They were in different departments with different collaborators, and yet they both jumped on this bandwagon at roughly the same time to start this movement in public health around 1995-1997. As co-authors on her first paper, Pat had 2 statisticians: Mei-Cheng Wang from Hopkins biostats and Xiaonan Xue from biostats at SUNY Albany. Neither one is closely associated with multilevel modeling. Ana had NO statisticians on her first paper, but she did have Pat's husband (Carles), and the famously crazy Eyal Shahar. How they landed on this topic at roughly the same time is an interesting mystery in intellectual history. Carles seems one potential connection.

Finally, check out this 1990 document attached comparing available software, including Goldstein's MI and Raudenbush's HLM. 1990!!! The focus is on educational research. Note that there is no mention of Stata. It already existed in 1990, but did not introduce the -xtreg- commands for random effects models until version 4.0 in 1995. SAS had random effects linear models sooner, with PROC MIXED in 1992 and followed much later with non-linear models via GLIMMIX in 2005.

- JK

From: Sam Harper <sam.harper@mcgill.ca>
Sent: Monday, August 22, 2022 5:11 PM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>; Basile Chaix <basile.chaix@iplesp.upmc.fr>
Cc: Patricia O'Campo <Patricia.OCampo@unityhealth.to>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

I'm loving this discussion and completely agree, Jay, that a retrospective SER session would be fantastic. And I also tend to agree that the papers you listed starting around 1996/7 are probably most emblematic of the interest in epi (but probably need to add the Sampson 1997 Science paper on collective efficacy, since I think that was a big deal at the time).

I also dug out my ***master's*** thesis on IPV (terrible), which was essentially co-opting Pat's seminal work (thank you!), which also had a few I had forgotten about. Since it isn't strictly epi or published in epi journals, you may consider it just ancillary, but there certainly was a strong interest in multilevel analysis in the criminology literature in the early 1990s. Some of these early papers (e.g., Sampson 1987, Smith/Jarjoura 1989, Miethe 1993) just ignored clustering entirely but put in contextual variables, but I think Rountree 1994 and thereafter also used multilevel logistic. I'm curious about the editor at Social Forces around this time, since they seemed to publish a few of these.

- Sampson, R.J., Wooldredge, J.D. Linking the micro- and macro-level dimensions of lifestyle-routine activity and opportunity models of predatory victimization (1987) Journal of Quantitative Criminology, 3 (4), pp. 371-393.

- Smith, D.A., Jarjoura, G.R. Household characteristics, neighborhood composition and victimization risk (1989) Social Forces, 68 (2), pp. 621-640.

- Miethe, T.D., McDowall, D. Contextual effects in models of criminal victimization (1993) Social Forces, 71 (3), pp. 741-759.

- ROUNTREE, P.W., LAND, K.C., MIETHE, T.D. MACRO-MICRO INTEGRATION IN THE STUDY OF VICTIMIZATION: A HIERARCHICAL LOGISTIC MODEL ANALYSIS ACROSS SEATTLE NEIGHBORHOODS (1994) Criminology, 32 (3), pp. 387-414.

- Rountree, P.W., Land, K.C. Burglary victimization, perceptions of crime risk, and routine activities: A multilevel analysis across Seattle neighborhoods and census tracts (1996) Journal of Research in Crime and Delinquency, 33 (2), pp. 147-180.

Also found a few earlier papers that could be of interest (more epi/public health):

- Love, R.R., Brown, R.L., Davis, J.E., Baumann, L.J., Fontana, S.A., Sanner, L.A. Frequency and Determinants of Screening for Breast Cancer in Primary Care Group Practice (1993) Archives of Internal Medicine, 153 (18), pp. 2113-2117.

- Gibbons, R.D., Hedeker, D., Elkin, I., Waternaux, C., Kraemer, H.C., Greenhouse, J.B., Shea, M.T., Imber, S.D., Sotsky, S.M., Watkins, J.T. Some Conceptual and Statistical Issues in Analysis of Longitudinal Psychiatric Data: Application to the NIMH Treatment of Depression Collaborative Research Program Dataset (1993) Archives of General Psychiatry, 50 (9), pp. 739-750.

- Hedeker, D., McMahon, S.D., Jason, L.A., Salina, D. Analysis of clustered data in community psychology: With an example from a worksite smoking cessation project (1994) American Journal of Community Psychology, 22 (5), pp. 595-615.

- Jones, K., Duncan, C. Individuals and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework (1995) Health and Place, 1 (1), pp. 27-40.

- Congdon, P. The Impact of Area Context on Long Term Illness and Premature Mortality: An Illustration of Multi-level Analysis (1995) Regional Studies, 29 (4), pp. 327-344.

- Leyland, A.H. Examining the relationship between length of stay and readmission rates for selected diagnoses in Scottish hospitals (1995) Mathematical Medicine and Biology, 12 (3-4), pp. 175-184.

Sam

From: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Date: Monday, August 22, 2022 at 10:03 AM
To: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Cc: Patricia O'Campo <Patricia.OCampo@unityhealth.to>, Sam Harper <sam.harper@mcgill.ca>, Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: RE: multilevel modeling seems to be having a little resurgence these days

Goldstein tribute attached.

It gives some history of his ML software program:

A key academic partnership developed in the late 1980s when Jon Rasbash joined Harvey at the IoE. Jon had been working with Michael Healy at the London School of Hygiene and Tropical Medicine and was an excellent programmer. Between them Jon and Harvey developed a series of (at the time DOS-based) computer programs to fit more and more advanced multilevel models. ML2 was developed in 1988 and fitted 2 level models, followed by ML3 in 1990 for 3 level models and then finally MLN in 1995 in the same year as the RSS meeting mentioned in the introduction.

- JK

From: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Sent: Monday, August 22, 2022 10:49 AM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Patricia O'Campo <Patricia.OCampo@unityhealth.to>; Sam Harper <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

Sure machines can do a good work if they are well trained. We have been using satellites data for long now, which are another machine!

Goldstein applications if I am not wrong were related to school, classroom, and pupils data. So it is clearly out what started around 1995 in our field, which was related to area effects. Another condition of this literature, as important as ML models, was that census or other area data had become available in a digital format.

In Paris in 2001, I remember that 1990 census data were available in digital format but that I had to manually enter from a paper book 1982 census data.

So there would be a statistical analysis condition and a measurement condition to this literature.

Regards,
Basile

De: "Jay Kaufman, Dr." <jay.kaufman@mcgill.ca>
À: "Basile Chaix" <basile.chaix@iplesp.upmc.fr>
Cc: "Patricia O'Campo" <Patricia.OCampo@unityhealth.to>, "Sam Harper, Dr."

<sam.harper@mcgill.ca>, "Seungmi Yang" <seungmi.yang@mcgill.ca>
Envoyé: Dimanche 21 Août 2022 13:51:49
Objet: RE: multilevel modeling seems to be having a little resurgence these days

Letting machines take on that contextual characterization now……

From: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Sent: Saturday, August 20, 2022 12:22 PM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Patricia O'Campo <Patricia.OCampo@unityhealth.to>; Sam Harper <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

By no way I wanted to mean that Pat is not a pioneer obviously :)
And of course it is always useful to reiterate on issues that are not universally understood.
Moreover, it is always useful to clarify things from an epistemological and history of science perspective.
About papers using GEE, this is why I am very much attached to the notion of contextual analysis, and think it is reductionist to refer to the stream of studies only from the perspective of multilevel analysis. It suggests that the novelty was mostly about modeling while in fact it was also related to novel measurement opportunities. Around the 2000s, scholars started to get rid of administrative boundaries for the measurement of contextual factors, typically in physical activity and food environment studies, and it was already moving beyond the most caricatural ML studies where everything had to be strictly hierarchical.
Basile

De: "Jay Kaufman, Dr." <jay.kaufman@mcgill.ca>
À: "Patricia O'Campo" <Patricia.OCampo@unityhealth.to>, "Basile Chaix" <basile.chaix@iplesp.upmc.fr>
Cc: "Sam Harper, Dr." <sam.harper@mcgill.ca>, "Seungmi Yang" <seungmi.yang@mcgill.ca>
Envoyé: Samedi 20 Août 2022 07:36:40
Objet: RE: multilevel modeling seems to be having a little resurgence these days

Pat's contribution is definitely pioneering. But I consider the paradigmatic social epidemiology multilevel model to be a random effects model, and Pat's 1995 AJPH paper uses GEE. It does include variables from neighborhood and individual levels, but GEE treats this hierarchy as a nuisance and does not provide inference on, for example, variance partition components. I consider Pat's 1997 paper to be much more in line with what I consider the real multilevel social epidemiology movement that took off thereafter, which is why I believe we are now at the 25-year mark. I would mark the foundation of this movement with these seminal papers:

O'Campo P, Xue X, Wang MC, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. American journal of public health. 1997 Jul;87(7):1113-8. [670 citations]

Diez-Roux AV, Nieto FJ, Muntaner C, Tyroler HA, Comstock GW, Shahar E, Cooper LS, Watson RL, Szklo M. Neighborhood environments and coronary heart disease: a multilevel analysis. American journal of epidemiology. 1997 Jul 1;146(1):48-63. [1000 citations]

Singer JD. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of educational and behavioral statistics. 1998 Dec;23(4):323-55. [3700 citations]

Kreft IG, De Leeuw J. Introducing multilevel modeling. Sage; 1998 [4675 citations]

Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. Social science & medicine. 1998 Jan 1;46(1):97-117. [900 citations]

Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. American journal of public health. 1998 Feb;88(2):216-22. [1675 citations]

Kennedy BP, Kawachi I, Glass R, Prothrow-Stith D. Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. Bmj. 1998 Oct 3;317(7163):917-21. [1000 citations]

Basile notes two papers that I didn't catch from 1995 and 1996 (note that Bryk and Raudenbush book came out in 1992).
These are not as heavily cited as the ones I listed above, so I don't consider them as influential, but they do involve
the kinds of random effects models I am referring to.

Curiously, the Kleinschmidt paper uses a 2-level logistic model, for which it defines normal level-1 errors, which doesn't
really make sense. I don't think that model specification is actually valid. They don't say anything about software or
implementation. It looks like maybe they simply fit the logits in 2-level linear model with least squares. Carr-Hill is also
vague on the software implementation and the details on the model, but it seems to be a 2-level negative binomial, which
is challenging to fit even today. They thank Harvey Goldstein for assistance, but I am also suspicious of this specification.
While I agree with Basile that these are earlier by a year or two, I think that the quirkiness of the applications made these
less influential, and there may have also been a European vs North American divide, with the latter being more influenced
by Raudenbush and the former by Goldstein. There is also the issue that these authors were not primarily social epidemiologists.
Carr-Hill was a statistician and Immo Kleinschmidt was a South African focused on malaria. These papers for these
authors were sort of one-off curiosities, which contrasts with Pat, Ana, Ichiro, etc using multilevel models to build a new
social epidemiologic movement around the methodology.

- JK

From: Patricia O'Campo <Patricia.OCampo@unityhealth.to>
Sent: Saturday, August 20, 2022 1:44 AM
To: Basile Chaix <basile.chaix@iplesp.upmc.fr>; Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Sam Harper <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>
Subject: RE: multilevel modeling seems to be having a little resurgence these days

My first ML paper was actually published in 1995 and my second one was 1997.
https://ajph.aphapublications.org/doi/10.2105/AJPH.85.8_Pt_1.1092

While I agree that there have been significant advances, I'm not sure that the lessons learned/applied re:ML are universal as I often get requests to review papers where those lessons are not apparent. A retrospective or discussion about where we're going to set intentionality is always a good idea. -Pat

---

From: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Sent: Friday, August 19, 2022 6:35 PM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Sam Harper, Dr. <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>; Patricia O'Campo <Patricia.OCampo@unityhealth.to>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

Hi Jay,
Below are two of the earliest references I am aware of:
- Kleinschmidt I, Hills M, Elliott P. Smoking behaviour can be predicted by neighbourhood deprivation measures. J Epidemiol Community Health 1995;49 Suppl 2:S72-S77.
- Carr-Hill RA, Rice N, Roland M. Socioeconomic determinants of rates of consultation in general practice based on fourth national morbidity survey of general practices. Bmj 1996;312:1008-12.

Also, my opinion is that many lessons from multilevel analysis and even more so contextual analysis have been learned and totally integrated in subsequent literature. There is a huge monthly literature that investigate the effects of any kind of contextual factor on individual health behavior or health outcome, and they systematically use random effect modeling. These recent studies have learned from their elders how to properly control for individual factors using DAG based reasoning, and although such adjustment factors are not called compositional factors, it is pretty much the same idea. The same applies to the notion of spatial scale of contextual factors, and various other methodological aspects of contextual analysis, which was already discussed in these old times. So overall I think that, although we do not now overemphasize the use of ML models or contextual analysis reasoning as we were doing in these times, many of these concepts and methods have become the ground on which we implicitly build more advanced studies.

Best regards,
Basile

De: "Jay Kaufman, Dr." <jay.kaufman@mcgill.ca>
À: "Basile Chaix" <basile.chaix@iplesp.upmc.fr>
Cc: "Sam Harper, Dr." <sam.harper@mcgill.ca>, "Seungmi Yang" <seungmi.yang@mcgill.ca>, "Basile Chaix" <chaix@u707.jussieu.fr>, "Dr. Patricia O'Campo" <o'campop@smh.ca>
Envoyé: Mercredi 17 Août 2022 10:21:05
Objet: RE: multilevel modeling seems to be having a little resurgence these days

Yes, I am thinking of a symposium for the 2023 SER meeting, a 25 year retrospective on the MLM revolution in social epid,
started by people like Pat O'Campo and Ana Diez-Roux in 1997, and then ramped up by Kawachi, Subramanian, Merlo,
Lynch, et al. I'll call it: "How multilevel social epidemiology MAIHDA comeback" or something like that. But the focus now
is completely different; note how all the promises of contextual causal inference have petered out. That is the interesting
part: the models are now focused on prediction and model selection instead of causal inference. Remember Susser's
"Chinese Boxes"? All that complex stuff about contextual versus compositional effects (e.g. Duncan Jones K, Moon G. Context,
composition and heterogeneity SSM 1998); who talks that way now?

- JK

PS – Basile: I am in France for the year; I hope we cam meet up.

From: Basile Chaix <basile.chaix@iplesp.upmc.fr>
Sent: Tuesday, August 16, 2022 9:23 PM
To: Jay Kaufman, Dr. <jay.kaufman@mcgill.ca>
Cc: Sam Harper <sam.harper@mcgill.ca>; Seungmi Yang <seungmi.yang@mcgill.ca>; Basile Chaix <chaix@u707.jussieu.fr>; Dr. Patricia O'Campo <o'campop@smh.ca>
Subject: Re: multilevel modeling seems to be having a little resurgence these days

Thanks Jay.
ML analysis would be revitalized by the intersectional perspective.
Get ready for another 10 years of ML analyses then!
Basile

Envoyé de mon iPhone

> Le 6 août 2022 à 17:38, Jay Kaufman, Dr. <jay.kaufman@mcgill.ca> a écrit :
>
> Random effects neighborhood studies were all the rage in the 1990s. Then they kind of died out in social epidemiology, and now I
> see a bit of a come-back. Andrew Bell seems really good. - JK
>
> Jay S. Kaufman, Ph.D
> Department of Epidemiology, Biostatistics, and Occupational Health
> School of Population and Global Health
> Faculty of Medicine and Health Sciences, McGill University
> Suite 1200, 2001 McGill College Avenue
> Montréal, Québec H3A 1G1

email: jay.kaufman@mcgill.ca
website: www.jayskaufman.com
Phone: 514-398-7341
FAX: 514-398-4503